MAKING METAPHYSICS MATTER:
ESSAYS ON REASONS AND PERSONS

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by
Rebecca Chan

______________________________
Michael Rea, Director

Graduate Program in Philosophy
Notre Dame, Indiana
June 2017
© Copyright by
Rebecca Chan
2017
All Rights Reserved
MAKING METAPHYSICS MATTER:
ESSAYS ON REASONS AND PERSONS

Abstract
by
Rebecca Chan

Metaphysicians commonly encounter challenges to why metaphysics matters. Sometimes, the challenge arises because metaphysical disputes are thought to be unresolvable. If (as is often assumed to be the case) the natural sciences and theoretical considerations constitute the totality of our evidence, there will be many cases in which we are not epistemically justified in believing any one of several competing metaphysical theses. If metaphysical inquiry doesn’t guide us to truth, then it seems that the inquiry is futile and pointless. Other times, the challenge arises because even if we can figure out which theses are true, their truth is thought to be practically irrelevant for our everyday lives. I appreciate the beauty of a statue regardless of whether it is identical to or distinct from the clay that constitutes it. I love my cat regardless of whether there are one or many cats occupying nearly the same spatial region. I value the ways in which friendships have changed my life for the better regardless of whether I persist through these changes by perduring or enduring over time. If these challenges of unresolvability and practical irrelevance are insurmountable, then the practice of metaphysics is a tragic pursuit hopelessly cut off from truth and devoid of practical significance.

In large part, my dissertation is a response to these challenges. The four independent essays that compose it are united by a central theme: to what extent
does normative data roughly, facts about value and what we ought to do bear on metaphysical disputes? Approaching metaphysical disputes in light of normative data is significant in two ways. First, an ethics-first metaphysics methodology is epistemically significant because it provides a way to resolve some metaphysical disputes. Second, focusing on the connection between metaphysics and normativity shows the ways in which metaphysical concerns matter for our practical lives. Thus, the central theme of this dissertation provides a way to answer both the unresolvability and practical irrelevance challenges for metaphysics.
# CONTENTS

**ACKNOWLEDGMENTS** ....................................................... iv

**INTRODUCTION** ............................................................. 1

**ESSAY 1: ETHICS-FIRST METAPHYSICS** ............................... 7

1.1 Introduction ............................................................ 7

1.2 Metaphysical Disputes ............................................... 9

1.3 Metaphysical Epistemicism .......................................... 13

1.4 Normative Data ....................................................... 18

1.5 Five Objections and Replies ....................................... 22

1.5.1 What About Metaphysical Intuitions? ......................... 22

1.5.2 Why Not Other Data? ............................................ 24

1.5.3 Isn’t Metaphysics More Fundamental than Ethics? ....... 27

1.5.4 Supplemental Principles ....................................... 28

1.5.5 Does Normative Data Presuppose Moral Realism? ....... 30

1.6 Ethics-First Metaphysics ............................................. 32

1.7 Persons ................................................................. 34

1.7.1 Three Problems ................................................... 35

1.7.2 Two Questions .................................................... 38

1.7.2.1 How Many Persons? .......................................... 39

1.7.2.2 Which Person? ................................................ 41

1.8 Conclusion ............................................................ 48

1.9 Cited Works ............................................................ 49

**ESSAY 2: REVIVING THE MODAL ACCOUNT OF ESSENCE** .......... 51

2.1 Introduction ............................................................ 51

2.2 Two Conceptions of Essence ....................................... 54

2.3 The (Alleged) Virtues of Definition ............................... 56

2.3.1 Seemingly Inessential but Necessary Properties ....... 56

2.3.2 Virtues of Exclusivity .......................................... 59

2.4 A Methodological Aside ............................................. 61

2.5 The Revised Modal Account ......................................... 64

2.5.1 Stage 1: Determinateness ....................................... 64

2.5.2 Stage 2: Sparseness .............................................. 68

2.5.3 Summing Up ........................................................ 71
I am forever thankful to all of those who helped me complete this dissertation. Whatever good-making qualities it has, it has in virtue of them. By expressing gratitude to specific people in what follows, I fear that I have forgotten to acknowledge some who guided me along the way. To those people, I apologize, and hope that they will be as gracious in forgiving me for my forgetfulness as they were gracious in helping me get to this point.

First and foremost, I owe an unpayable debt of gratitude to Michael Rea, my advisor. Over the last few years, Mike has read far too many drafts and suffered through my half-baked ideas, poorly executed paper attempts, and grammatically deficient writing. Through it all, he never failed to encourage me to pursue these projects and always showed me how to make my work better. Perhaps even more significant than his philosophical guidance has been Mike’s support, generosity, and kindness: he brought me to the Center, engineered my transfer to a metaphysics haven, funded conferences, guided me through the job market, let me occupy his office, and was simply always there with advice and encouragement. I have never been left wanting. Mike has shown me how to be a good philosopher and reminded me to be an even better person. I am extremely fortunate to have as an advisor someone to whom I can aspire to emulate.

Second, I’d like to thank the other members of my committee: Christopher Shields, Meghan Sullivan, Peter van Inwagen, and L. A. Paul. I had almost forgotten how star-struck I was by each of them until the defense, when I felt the full force of their intellect gathered in one room. But in addition to being...
philosophical titans, the four of them proved to be incredibly giving with their feedback. They challenged the weakest points of my arguments, gave me constructive advice on how to make them stronger, and upped my philosophical game, so to speak. My dissertation is as good as it is because of them, and I am thankful to them for their guidance throughout the process.

Third, I am tremendously appreciative to so many for their comments and discussion that helped in the development this project. Robert Audi, David Barnett, Sara Bernstein, Andrew Brenner, Anjan Chakravartty, Eric Chwang, Dustin Crummett, Trent Dougherty, Peter Finocchiaro, Jason Hanschmann, Jack Himelright, Liz Jackson, Anne Jeffrey, Kathrin Koslicki, Ting Cho Lau, Christian Lee, Michael Longenecker, Michaela McSweeney, Bradley Monton, Graham Oddie, David Patillo, Callie Phillips, Mark Puestohl, Ben Rohrs, Meg Schmitt, Jesse Schupack, and the Metaphysics Reading Group all read parts of the dissertation and provided feedback. My year in the Ontology Room with Peter, Callie, David, Michael, Andrew, Jack, and Dustin was especially beneficial—I learned much from them and thoroughly enjoyed our many conversations. I must also mention Wes Morriston and Garth Green, the two people who sparked my love of philosophy. Without them, I would have never aspired to write a dissertation and would be living an intellectually impoverished life.

Fourth, I am especially thankful to Jason and Michaela who have been my friends and interlocutors since I started my career as a graduate student years ago in Colorado. In addition to always being available to read horrible drafts on short notice and talk through ill-formed ideas, they’ve tolerated years of neuroticism and managed to keep me sane (or at least functional). I hope that we have many more years of philosophy, climbing, and adventure ahead of us.

Fifth, I would like to express my gratitude and love for the other members of Team R.E.A., Peter and Callie. They welcomed and supported me from the moment
I joined their ranks, and I would not have been able to write as well as I did without them. I could not have hoped for better philosophy siblings, and I hope that I have been as good a teammate to them as they have been to me.

Finally, I am thankful for Ponyo, who faithfully sat with me through long days and late nights as I started, and eventually finished, my dissertation.
INTRODUCTION

Metaphysicians commonly encounter challenges to why metaphysics matters. Sometimes, the challenge arises because metaphysical disputes are thought to be unresolvable. If (as is often assumed to be the case) the natural sciences and theoretical considerations constitute the totality of our evidence, there will be many cases in which we are not epistemically justified in believing any one of several competing metaphysical theses. If metaphysical inquiry doesn’t guide us to truth, then it seems that the inquiry is futile and pointless. Other times, the challenge arises because even if we can figure out which theses are true, their truth is thought to be practically irrelevant for our everyday lives. I appreciate the beauty of a statue regardless of whether it is identical to or distinct from the clay that constitutes it. I love my cat regardless of whether there are one or many cats occupying nearly the same spatial region. I value the ways in which friendships have changed my life for the better regardless of whether I persist through these changes by perduring or enduring over time.

If these challenges of unresolvability and practical irrelevance are insurmountable, then the practice of metaphysics is a tragic pursuit hopelessly cut off from truth and devoid of practical significance. In large part, my dissertation is a response to these challenges. The four independent essays that compose it are united by a central theme: to what extent does normative data—roughly, facts about value and what we ought to do—bear on metaphysical disputes? I won’t argue that the answer to this question is the same for all metaphysical disputes. Instead, different answers will be appropriate for different disputes. Sometimes, normative data won’t bear
on the metaphysical dispute at all. In these cases, it will either be the case that the metaphysics is completely independent of normative concerns (independence), or that the metaphysics is prior to the normative concerns (metaphysics-first ethics). Other times, the normative data will bear on the metaphysical dispute. Here, there are different degrees to which the normative data bears on the metaphysical dispute. At the weakest end of the spectrum, normative concerns might tell us which entities are of special practical concern (practical ethics-first metaphysics). (Practical ethics-first metaphysics appears only in passing in the essays, though it is entailed by each of the positions on the stronger part of the spectrum.) Slightly stronger would be a dispute in which normative data epistemically bears on the truth of metaphysical theses (ethics-first metaphysics). Finally, at the extreme end of the spectrum, normative facts might determine metaphysical ones (determining ethics-first metaphysics).

Understanding these ways in which normative data bears on metaphysics is significant in two ways. First, an ethics-first metaphysics methodology is epistemically significant because it provides a way to resolve some metaphysical disputes. Second, both the metaphysics-first and ethics-first approaches show the ways in which metaphysical concerns matter for our practical lives. Thus, the central theme of this dissertation provides a way to answer both the unresolvability and practical irrelevance challenges for metaphysics.

**Essay 1: Ethics-first Metaphysics**

In my first essay, I set the stage by developing the ethics-first metaphysics methodology. On the ethics-first approach to metaphysics, we begin by surveying normative data. Epistemically speaking, normative data is reliable. In addition, normative facts and metaphysical theses are often necessarily connected. For instance, it is a piece of normative data that sometimes, some agents are morally responsible for their actions. Plausibly (and perhaps analytically) an agent’s having
free will is a necessary prerequisite for their being morally responsible. Thus, there are instances of free will. From a piece of normative data, we arrive at an interesting metaphysical thesis.

Of course, this ethics-first approach invites many questions, so I spend the bulk of the chapter motivating the approach, defending it from some objections, and explaining its scope. In examining what the scope of this approach is, this essay also answers the question of when metaphysical disputes display independence from normative concerns. I ultimately conclude that disputes about persons and reasons are the ones upon which normative data bears. This conclusion makes sense given that persons are the subjects (and sometimes objects) of normativity and that reasons carry normative force. The essay ends by considering an application of the ethics-first metaphysics methodology to persons.

**Essay 2: Reviving the Modal Account of Essence**

The remaining three essays focus on persons, the subjects (and, sometimes, objects) of normativity, and reasons, the things that carry normative force. My second essay engages with persons' metaphysical profiles, or what are typically called essences. On a popular conception of essence, a thing’s essential properties are the ones that define what it is. I argue against this definitional conception, defending instead a modal conception of essence on which all of thing’s necessary properties are essential to it. Though the modal account as traditionally formulated runs into problematic counterexamples, I offer two modifications, one based on determinacy and one based on sparseness, that avoid these counterexamples while preserving the primary virtues of the traditional account.

It may initially appear that this essay is the outlier of the four: the dispute about how best to characterize essence appears to be one that has independence from normative considerations. However, this appearance is misleading, for this
essay contains two interesting implications for how a *metaphysics-first ethics* might go. First, metaphysical profiles turn out to be important because they limit the candidate practical profiles of a person. A person’s practical profile cannot lack an essential property from its metaphysical profile, though it can have additional essential properties. This point surfaces in the discussion of person-candidates in the first and third essays. In these discussions, person-candidates are characterized modally, and it is unclear that a definitional characterization would adequately capture the differences between the person-candidates. Thus, we have a case in which a seemingly pure metaphysical issue provides a framework in which to understand a person’s practical profile.

Second, I propose that essentiality can admit of degrees and that one of the modifications to the traditional modal account is that the degree to which a necessary property is essential is proportionate to the degree to which it is determinate. A property is more (or less) determinate in proportion to how narrow (or wide) the scope of what a thing that has that property could be. Practical profiles are more narrow than metaphysical ones because they contain additional essential properties. For instance, the property of being a philosopher traditionally is not considered part of any person’s metaphysical profile. However, it could very well be part of a person’s practical profile. Thus, the property of being a philosopher is relatively determinate, provided that it is an essential to some practical profile. If this is right, then there is a surprising practical upshot of this metaphysical discussion about the best way to characterize essences. The distinctive, necessary properties like being a philosopher that make up a person’s practical profile are amongst that person’s most essential properties. Who we are as practical agents turns out to be rather essential to what we are metaphysically speaking, and the correct view of metaphysical essences explains why this is.
Essay 3: The Problem of Self-Transformation

Practical profiles, which I turn to in my third essay, differ from metaphysical ones by including features that we consider essential to our agentive lives such as our core preferences and life goals. We might think, in accordance with the orthodox view in decision theory, that when deliberating on which potential future best satisfies self-interest, one should first narrow the options to those futures that contain my metaphysical profile then see which of those involves the most value for oneself. I argue that cases involving radical transformation show that this procedure is misguided because it fails to recognize that people also have practical profiles, and that these practical profiles are what matter for self-interest.

In this essay, I also propose that persons, to a certain degree, can shape their practical profile. This discussion goes beyond the discussion of persons in the first essay. Normative concerns don’t just epistemically lead us to which person candidate is the person; in addition, it turns out that we can create normative facts about our self-interest that then determine which person candidate is the person. Thus, this essay presents the extreme end of the spectrum involving determining ethics-first metaphysics.

Essay 4: Religious Experience, Voluntarist Reasons, and the Transformative Experience Puzzle

Finally, in the fourth essay, I explore whether there might be voluntarist reasons, or reasons originating from an act of the will. Reasons are usually taken to be either mind-independent entities or psychological states, so whether reasons can originate from the will is controversial. However, if there are such reasons, they allow agents to respond to a puzzle posed for decision theory by transformative experience. Transformative experiences involve inscrutability with respect to what an experience will be like and how having the experience will change an agent, so
agents cannot decide to undergo these experiences based on what they project the
experience will be like. Voluntarist reasons are practically significant in part
because they enable agents to make rational decisions in these cases.
Abstract. This paper is about normative data and the work they can do. Normative data, roughly, are facts about value and what we ought to do. Besides informing our actions, they also can inform us about the way the world is. This latter work is the focus of this paper. Let *metaphysical epistemicism* be the view that metaphysical disputes are meaningful (i.e., not merely verbal) but irresolvable because there is little justification for believing either side of the dispute. This view is troubling because (i) it seems true if data given by the sciences and theoretical considerations constitute the totality of our evidence with respect to metaphysics theses, and (ii) given its truth, the project of discovering which metaphysical theses are true is futile. I argue that normative data are an overlooked source of evidence. Thus, even if (i) is true, we are not forced into metaphysical epistemicism and can make progress on metaphysical disputes, thus avoiding (ii).

1.1 Introduction

David and Peter are engaged in a dispute about tables. Specifically, they disagree about whether tables exist. David shares the common-sense notion that many of us have—that tables exist. Sometimes, simples (say, indivisible, point-sized particles) are arranged such that there is an elevated flat surface upon which other items can be placed. In these situations, those simples compose a new object that is something over and above the collection of simples, namely, a table. Peter does not deny that there are simples arranged table-wise; on this point David and Peter agree. But Peter does deny that there is a table that the simples compose. According to Peter, there are merely simples arranged table-wise, and no further object that is composed by these simples.
This dispute strikes many as peculiar, and metaphysicians, especially in the recent metametaphysics literature, have acknowledged this peculiarity and attempted to diagnose it. Perhaps David and Peter are using ‘exists’ in different ways and are engaged in a merely verbal dispute. Or perhaps existence questions are either trivial or meaningless, in which case the dispute is strange because the answer is obvious or inapplicable. Alternatively, the dispute might be substantive, non-trivial, and meaningful, but strange because we do not have the means to resolve the dispute; the available evidence is insufficient for epistemically settling the dispute. Finally—and this diagnosis is more commonly offered by non-metaphysicians—one might think that the dispute is peculiar because it does not matter. Nothing of practical concern rests upon whether David or Peter is right.

In this paper, I’m going to assume that metaphysical disputes are substantive, non-trivial, and meaningful. They can be resolved, in principle, though I leave as an open question whether we have the epistemic means to resolve the dispute. Borrowing from Bennett (2009), let metaphysical epistemicism be the view that metaphysical disputes are meaningful, or not merely verbal, but irresolvable because there is little justification for believing either side of the dispute. The goal of this paper is to show that metaphysical epistemicism is too pessimistic. We can make progress on some, though not all, metaphysical disputes by appealing to evidence given to us by the normative domain. Interestingly, this way of resolving disputes also offers a response to the criticism that metaphysical disputes do not matter.

---

1 Hirsch (2009), for instance, argues for this diagnosis while others like Sider (2009) argue against him.

2 For instance, Schaffer (2009) argues for permissivism about existence while Thomasson (2009) argues that the existence of controversial entities follows trivially from language. While their views differ in significant ways (Thomasson’s existence is more deflationary than Schaffer’s), both find existence (too) easily resolved. There’s also the Carnapian response to existence questions, on which these questions are meaningless.

3 E.g., Dennett (2006); Parfit (1984); van Fraassen (2002).
Here’s the plan for the rest of the paper. I start by getting clear on what
metaphysical disputes are (§ 1.2). Next, I turn to metaphysical epistemicism and
discuss why it arises when the totality of our evidence comes from the natural
sciences and theoretical considerations (§ 1.3). I then make a positive proposal: for
some metaphysical disputes, we can avoid epistemicism by turning to an overlooked
evidential source—normative data, or facts about value and what we ought to do
§ 1.4). After getting the positive proposal on the table, I consider and respond to
five potential worries (§ 1.5). The above considerations lead us to an “ethics-first
metaphysics” methodology (§ 1.6). I then explore an application of this approach to
one of the paradigmatically “unresolvable” disputes, that of coincident objects
§ 1.7). Finally, I offer some concluding remarks on the project (§ 1.8).

1.2 Metaphysical Disputes

Let a metaphysical dispute be a disagreement in which competing, incompatible
metaphysical theses are advanced. I take a metaphysical thesis to be a proposition,
or conjunction of propositions, where the content of the proposition(s) concern the
subject matter of metaphysics. Characterizing the subject matter of metaphysics
is difficult, but for now let’s use a suggestion offered by [Paul] (2012). According
to Paul, the subject matter of metaphysics is continuous with that of the natural
sciences. Both use the same methodology, and the boundary between the two falls
roughly between the empirical and non-empirical. Presumably, scientific theories are
distinguished primarily by their empirical predictions, so science concerns what is
subject to empirical observation while metaphysics does not, at least not directly.
Paradigm scientific issues include what the world is like at the molecular level, while
paradigm metaphysical issues include how many ontological categories there are.
On this understanding, David and Peter’s disagreement is a paradigm example of a
dispute concerning the subject matter of metaphysics. The two advance incompatible
theses—that simples arranged table-wise compose a further object and that it is not the case that the simples arranged table-wise compose a further object—that are about the subject matter of metaphysics.

This characterization of the subject matter of metaphysics is not completely unproblematic. First, one might question whether the characterization is apt. There are some scientific theses (e.g., whether there is absolute velocity and position in addition to relative velocity and position) that are empirically equivalent. Some metaphysical theses (e.g., whether there can be genuinely indeterminate events) may fall within the purview of scientific discovery. For instance, Balaguer (2010) argues that empirical facts bear on whether there is libertarian free will. Other metaphysicians have argued that considerations from science such as special relativity and Newtonian laws bear on debates about time and composition, respectively. But taking a closer look at these cases suggests that Paul is right about the boundary. In addition to holding that empirical facts bear on the libertarian hypothesis, Balaguer also holds that the hypothesis is an empirical rather than metaphysical one. What may have initially appeared to be an instance of an empirical metaphysical hypothesis turns out to respect the proposed boundary. Furthermore, in the arguments regarding time and composition, the scientific considerations that are brought to bear on metaphysics tend to be relatively theoretical. The considerations also do not appear to have settled the disputes.

Perhaps science eliminates potential metaphysical theses that are incompatible with its empirical discoveries, but these are not the theses that are the subject of live metaphysical debates. This suggests that the border of science and

---

4 For instance, one argument against presentism is that presentism cannot make sense of some of the predictions of special relativity. Bigelow and Pargetter (2006) present an argument against nihilism about composition that uses considerations from Galileo’s thought experiment that shows that objects fall at the same rate regardless of their mass.

5 Sociologically speaking, there are still many presentists and nihilists. They also appear to be both rational and aware of the arguments.
metaphysics is located around the line between the empirical and non-empirical, though the boundary might be vague. If we accept this distinction between science and metaphysics, then most paradigm metaphysical disputes aren’t going to be resolved by anything likely revealed by the sciences.

Second, one might worry that the characterization given by Paul is not apt because of cases where evidence for metaphysical claims clearly fall on the empirical side of the boundary. For example, we know that there are filled table-shaped regions, regardless of whether those regions are filled with simples arranged table-wise or tables. Perhaps this claim is open for debate in the context of skeptical debates in epistemology, but metaphysically speaking, there is virtually no debate with respect to whether that claim is true. Debates, like David and Peter’s, take place downstream of agreement on that claim. Even idealists can accept that claim and then have a debate over whether those simples arranged table-wise are ideas or physical objects.

In addition to these uncontroversial existential claims, there are also related uncontroversial modal claims that we know. If we know that there are filled table-shaped regions of space, then we know that possibly, there are filled table-shaped regions of space. More ordinarily in everyday life, we make modal claims such as “I can pick you up from the airport,” or “Possibly, my cat retrieves trash”; presumably, we know that these claims are true from experience. If all modal claims are metaphysical ones, then there are metaphysical claims that we know in virtue of empirical observation.

There are a couple ways in which one might respond to this worry. First, one might try to deny that these existential and modal claims are genuine metaphysical ones. After all, anyone with basic observational capacities can know that there are filled table-shaped regions of space or that possibly, there are filled table-shaped regions of space. A second response may be to simply acknowledge that in these cases, claims that concern the subject matter of metaphysics can be based on the
empirical. We just don’t typically consider these claims when thinking about the subject matter of metaphysics because they are not the claims under dispute. Our knowledge of these claims is justified by observation (e.g., that there is a filled table-shaped region) and conceptual analysis (e.g., that actuality entails possibility). This knowledge is as secure as our knowledge of the natural sciences, and even those who are most skeptical of our metaphysical knowledge don’t deny that we can know these sorts of metaphysical truths. The subject matter of metaphysical disputes lies on the non-empirical side of the boundary even if there are some metaphysical claims that do not. In a way, this makes sense, since work is done where there are disputes rather than where there is widespread agreement because the issue has already been settled by readily available empirical evidence.

Before moving on to the next section, it’s worth highlighting two points from this discussion. First, metaphysical claims that we can know are ones for which we have evidence from experience or conceptual analysis. The claims that are under dispute lack that evidence. This is significant for the next section. Second, the irresolvable disputes that are of interest for this paper are not in principle irresolvable. For instance, an omniscient being could know when, if ever, composition occurs. (If there are cases in which it’s indeterminate whether composition occurs, the omniscient being would know that as well.) And if an omniscient being were to enlighten us about the facts regarding composition and we were in a position to know that the being was omniscient, we would be in a position to know the facts about composition. Instead, these disputes are irresolvable because we are not omniscient and occupy a limited epistemic position. Let’s now turn to why our epistemic situation is too impoverished to resolve many of these disputes.

---

6 See van Inwagen (1998). There are uninteresting possibility claims that we can make on the basis of what is actual. But the interesting modal claims involve universal claims and claims that x is possible where x is not actual. These interesting claims are the ones of which we should be skeptical, according to van Inwagen.
1.3 Metaphysical Epistemicism

Suppose that David and Peter disagree about the physical, rather than metaphysical, composition of the table (or simples arranged table-wise). David thinks that the table is made out of granite while Peter thinks that the table is made out of cleverly disguised faux-granite. There is a straightforward way of resolving this dispute. They can take a sample of the table to a lab and run some tests that will determine what the physical make-up of the table is. There is available empirical evidence, i.e., evidence given by the natural sciences, that they can obtain that will settle the dispute.

Alternatively, they might appeal to a mix of empirical evidence and theoretical considerations. For instance, suppose that they discover that the table has been purchased from a store that has recently been involved in a peculiar scandal. The store has been passing off faux-granite tables as granite tables in order to increase their profit margins. Those who uncovered the scandal project that approximately 99% of the tables from this store are in fact faux-granite tables. In this scenario, an empirical fact combined with a rather uncontroversial inference would provide strong justification for Peter’s thesis. Perhaps the company selling the table sells them at shockingly low prices. The faux-granite thesis has the virtue of fruitfulness, since it explains how the company can afford selling at such low prices. Even though these pieces of evidence don’t deductively entail that Peter is correct, they do provide extremely strong support for his thesis.

Metaphysical disputes, such as ones over composition, differ from this physical composition dispute in two ways. First, while there is empirical evidence that can settle the physical dispute, there is no empirical evidence that can settle the metaphysical composition dispute. Second, theoretical considerations—specifically, theoretical virtues—sometimes help resolve physical disputes. I take the theoretical virtues to be content-neutral features of theories—such as consistency, parsimony,
and fruitfulness—that count in favor of adopting that theory. They do not help in metaphysical disputes, or at least not in the ones that are heavily contested, because competing theses tend to exhibit roughly equal degrees of theoretical virtue. I’ll elaborate on these two differences in a moment, but I want to first note that these differences, if they exist, frustrate the truth goal of metaphysics—the goal of discovering which metaphysical claims are true. Since data from the sciences and theoretical considerations are typically thought to exhaust the available evidence, if that evidence doesn’t bear on a metaphysical dispute, the dispute will be irresolvable. If the metaphysician wants to avoid this result, she must either deny one of the two evidential differences between metaphysics and natural science or hope for an additional source of evidence. I’ll spend the rest of this section explaining why the first option is not viable.

First, let’s consider whether empirical data given to us by the natural sciences bear on metaphysical disputes. Suppose that David and Peter know all the physical facts about the simples in front of them that are arranged table-wise. These facts are compatible with the truth of either thesis regarding composition. There is nothing they could learn about the physical makeup of the simples and their arrangement that would settle whether composition occurs or not. This difficulty applies to other metaphysical disputes as well. For instance, the physical facts about the table also fail to settle whether there might be objects that are co-located with the table. After all, those who hold that there can be coincident objects share claim that those objects share all of their categorical properties; the entirety of what there is to observe will not support a overlap thesis over a non-overlap thesis, and vice versa.

The competing theses in David and Peter’s dispute are empirically equivalent. Let two theses be empirically equivalent if and only if the truth (or falsity) of both is logically compatible with all available data given by the natural sciences. It may even be the case that these metaphysical theses are necessarily empirically
equivalent. In the case of the disputes we’ve considered thus far, there is *in principle* no empirical data we can obtain that will settle the disputes.\footnote{It may even be that metaphysical theses regarding composition are all compatible with scientific theory. Rosen and Dorr (2002) make this stronger point when they question what it would even be for a mereological thesis to conflict with a scientific theory.} In fact, metaphysically intractable disputes might turn out to be necessarily empirically equivalent in virtue of the subject matter of metaphysics. Recall from §1.2 that intractable disputes fall on the non-empirical side of the empirical-non-empirical line. Given that, we shouldn’t expect to find empirical data that bears on metaphysical disputes. Thus, the empirical data underdetermines which of the theses is true since the data is compatible with both. As a result, one major source of evidence fails to advance the truth goal by not providing justification for believing one competing thesis rather than the others.

A second source of evidence comes from theoretical considerations. This type of evidence can take two forms. First, conceptual analysis can lead us to truth. For instance, perhaps there’s a compelling definition of ‘part’ that would yield an analytic resolution to the debate over composition. There doesn’t appear to be such a definition, but if there were, when composition occurs would be a conceptual truth.\footnote{Rosen and Dorr (2002) make this point about the apparent hopelessness of using conceptual analysis to resolve disputes about constitution.} Besides giving us conceptual truths, conceptual analysis can yield non-conceptual truths when combined with other evidence. For instance, our earlier claim that possibly, there are simples arranged table-wise, is a combination of the observation that there are these simples and that actuality entails possibility. Of course, conceptual analysis only takes us so far. It can decisively justify the truth of a thesis, but that thesis won’t be one of the ones that is part of an intractable dispute, such as the one over composition.

The second type of theoretical consideration is more controversial. Competing
theses and theories might display theoretical virtues such as parsimony and fruitfulness. Having these virtues does not entail truth. Nevertheless, the virtues of a theory are often presented as truth-conducive. For instance, Lewis grounds his belief in modal realism in its fruitfulness, claiming that the work modal realism does “gives us good reason to believe that it is true” (Lewis 1984, 4). However, using these appeals to justify belief in a thesis is problematic. First, it is unclear that virtues such as parsimony or fruitfulness are truth-tracking. Only internal consistency, or rather a failure of a thesis or group of theses to be internally consistent, would yield a definitive verdict on truth. Presumably, metaphysical theses that lack internal consistency have already been ruled out and are not amongst the competitors.

Second, even if there are scenarios in which theoretical virtues might provide decisive evidence for a thesis, the metaphysical disputes with which we are concerned aren’t cases in which this approach will work. For instance, Bennett (2009) points out that in the universalism-nihilism debate about composition, competing theses do roughly as well as each other with respect to overall theoretical virtue. Though the nihilist might enjoy a slight advantage over the universalist with respect to ontological parsimony, the universalist might enjoy a slight advantage when it comes to ideological parsimony. On the whole, both positions do roughly as well as the other. Competing theses in intractable metaphysical disputes are rough equals with respect to theoretical virtues. Let two theses be theoretically equivalent if and only if they display roughly the same degree of theoretical virtue. Where competing theses are theoretically equivalent, there is no justification for believing one thesis rather than another on the basis of theoretical considerations. Thus, in addition to a lack of empirical grounds, there is a lack of theoretical grounds on which to settle these debates.

9See, for instance, Huemer (2009) for an argument against parsimony as truth-tracking.
In light of these two types of equivalence, resolving metaphysical disputes appears hopeless. For example, Bennett (2009) argues that some debates in metaphysics are unresolvable because we are not epistemically situated to answer questions about whether certain (types of) entities exist. For these reasons, Bennett argues that when we inquire as to whether there are Fs, such as tables, we ought to answer:

‘There are F’s’ is either true or false, and disputes about its truth-value are not verbal disputes. But there is little justification for believing either that it is true or that it is false. (5)

Like Bennett, Rosen and Dorr (2002) make a similar point about composition:

The truth about composition—whatever it may be—is neither analytic nor straightforwardly empirical...Our case for agnosticism about composition...is the upshot of having canvassed the main sources of grounds or evidence and come up wanting. (24)

Empirical and theoretical grounds are not sufficient for settling metaphysical disputes. Unfortunately, if those are the only two evidential grounds available to metaphysics, we won’t be justified in believing of any metaphysical thesis that it is true.

This result is disappointing. But even if we accept it, we might wonder if metaphysics still has an important role to play despite the fact that its disputes are unresolvable. Metaphysicians have suggested that there are still important projects associated with metaphysics. Paul (2012) points out that metaphysics is still useful for clarifying concepts and theses. Even if there are no grounds that justify belief in nihilism or universalism about composition, we can get clearer on what each position involves. In addition, metaphysicians excel at determining which metaphysical theses are compatible. Indeed, many interesting metaphysical questions have to do with which positions are consistent with each other. For example, we might wonder whether libertarian free will is compatible with an eternalist theory of time or divine foreknowledge. Let these two points be the

---

10 As Rosen and Dorr (2002) succinctly put it: “Would it be better to know? Of course it would.”
*consolation of metaphysics*. They leave valuable work for metaphysics, but fall short of the aspirations of metaphysicians who seek truth.

Is metaphysics forced to settle for this consolation? It may appear that it is given empirical and theoretical equivalence. However, settling in virtue of this reason is too hasty if empirical data and theoretical virtue are not the only sources of evidence that are relevant to the truth of metaphysical theses. As I suggest in the next section, normative data provides a third evidential source. Thus, the metaphysician seeking truth need not settle for the consolation of metaphysics quite yet.

1.4 Normative Data

Thus far, I’ve argued that empirical data and theoretical considerations underdetermine which competing thesis in a metaphysical dispute is the one that we are justified in believing. This result leads to metaphysical epistemicism if those two sources exhaust all of the evidence available to us. Some, like Bennett, accept this consequence and embraces metaphysical epistemicism. In this section, I propose that there is a third source of evidence that is relevant to metaphysics—normative data. I’ll start by explaining what normative data are and why we ought to regard them as an evidential source. I’ll then sketch how it might help us make progress with respect to some metaphysical disputes.

Normative data, as I understand it, consists of facts regarding value and what we ought to do. For our purposes, we can conceive of normative data broadly. For instance, it can encompass deontic facts, such as the fact that I ought to be kind to strangers or the fact that I ought to refrain from inflicting gratuitous harm. Normative data can also include axiological facts about value, such as the fact that pain is bad and beneficence good. It might also include what we ought to do in more broad or narrow senses. For example, the fact that I all things considered ought to donate a substantial percentage of my income is a piece of normative data,
as is the fact that from the standpoint of self-interest I ought to spend my income on making myself happy. Finally, normative data can be a priori or a posteriori. For instance, some of this data might come from intuitions, which are typically considered a priori. The data might also come from observation and thus be a posteriori; Oddie (2005), for example, holds that desires are perceptions of value. I’m inclined to think that normative data is varied and includes both axiological and deontological facts, global and self-interested perspectives, and a priori and a posteriori sources. But one need not be committed to this pluralistic view. All that matters for the present purposes is that there is some normative data.

Crucially, normative data, even on its broadest construal, are distinct from the two evidential sources, empirical data and theoretical considerations, discussed in the previous section. The empirical data given to us by the natural science is descriptive. It explains the physiology of pain, but doesn’t tell us that pain is bad or that it ought not be inflicted gratuitously. Similarly, evolutionary biology might offer a story on which beneficence is evolutionarily advantageous, but it doesn’t tell us that beneficence is good or that we ought to promote it. In short, the nature of empirical data is descriptive, while the nature of normative data is prescriptive and value-laden. For those inclined to think that ‘ought’ cannot be derived from ‘is’ it should be unsurprising that normative data is distinct from empirical data. But even if one is inclined to think that normative facts supervene on physical facts, the natural sciences clearly deal only with the physical and are silent on the normative and the relationship between the two domains.

Normative data is also distinct from the theoretical considerations that we have considered as an evidential source. It is not itself a theoretical virtue. That pain

---

11 I realize that observations of value are empirical, in the strict sense of that word. For the purposes of this paper, I’m reserving the term ‘empirical’ for referring to data from the natural sciences. ‘Normative data’ refers to facts about value and what we ought to do, some of which may be obtained via experience.
is bad and that we ought to avoid inflicting it are not a theoretical virtues the way parsimony or consistency are virtues of metaphysical theses. It may turn out that given a piece of normative data, theoretical considerations like consistency favor one metaphysical thesis over another. (In fact, that’s what I’m about to suggest!) But it’s significant that normative data is a distinct evidential source than theoretical considerations.

If normative data is a distinct, third evidential source, then there is still hope for the metaphysician who cares about truth. Even if theses are empirically and theoretically equivalent, competing theses may not be normatively equivalent. Let two theses be normatively equivalent when and only when the truth (or falsity) of both are compatible with all available normative data. I suspect that there is an unfortunate tendency to assume that metaphysical theses are normatively equivalent. (This assumption is perhaps what motivates Parfitian dismissiveness of metaphysics on the grounds that it does not matter.) It’s true that some metaphysical theses are normatively equivalent, just as some metaphysical theses are empirically and theoretically equivalent. For instance, whether simples arranged table-wise constitute a further object does not affect my obligation to respect your property rights; it is impermissible for me to steal things that you own, whether you own a mere collection of simples arranged table-wise or those simples and a table. Theses that differ in structure or explanatory direction might be normatively equivalent. Consider, for instance, the Euthyphro Dilemma. On one horn, an act is pious because the Gods approve of it. On the other, the Gods approve of an act because it is pious. These two horns contain very different theses: one has piety as more fundamental than the Gods’ approval and the other has Gods’ approval as more fundamental than piety. Nevertheless, on both horns, one ought to act piously.

Fortunately, not all metaphysical disputes involve competing theses that are normatively equivalent. In cases where the competing theses are not normatively
equivalent, normative data can help us make progress with respect to at least some disputes. Here’s a toy example of how this might go. Sometimes, we observe that someone is morally responsible for their actions. They’ve performed an act that they ought to have performed, or failed to perform one that they ought to have, and we observe that they have done this. For instance, we just see that some person is morally responsible for her act of kindness towards a stranger. That this person is morally responsible for her good action is a piece of normative data. Now plausibly, freedom is a necessary precondition for moral responsibility. This claim might even be true analytically. If this is right, there is a normative fact—that a person is responsible for her good action—that takes us to an interesting metaphysical thesis—that there is freedom.

Of course, taking normative data into account when settling metaphysical disputes may not rule out all but one competing thesis. Suppose that normative data establishes that there is freedom. This rules out theses that are incompatible with there being some instances of freedom, such as hard determinism:

**Hard Determinism:** All events are determined and determinism is incompatible with freedom.

But there are multiple theses that are compatible with the falsity of **Hard Determinism.** Consider the following two theses:

**Libertarianism:** An agent performs some act freely iff the agent could have chosen to do otherwise, and there are some instances of agents acting freely.

**Compatibilism:** An agent performs some act freely iff the agent chooses to perform that action, and there are some instances of agents acting freely.

One way of understanding the debate between **Libertarianism** and **Compatibilism** is that it is a disagreement on what the preconditions for freedom are. However one
comes down on that debate, without further evidence, the two theses are normatively equivalent. While the normative data rules out views on which there is no freedom, it does not distinguish between views that hold that there is freedom but differ on the preconditions of freedom.

What I’ve offered so far is a sketch of how normative data might allow for metaphysical progress. My claim is not that normative data can resolve all metaphysical disputes, or that it can definitively resolve disputes such that there is exactly one thesis from the competitors that we are justified in believing. Instead, I’m defending a rather modest position: there are competing metaphysical theses that are not normatively equivalent, and in these cases, normative data allows us to make progress with respect to which thesis we are justified in believing is true.

1.5 Five Objections and Replies

In the previous section, I offered a positive proposal for how normative data helps mitigate metaphysical epistemicism with respect to some disputes. However, there are worries one might have about this proposal. In this section, I’ll consider and address five objections to the view I have proposed. The first two relate to the issue of whether there is a special category of metaphysical data that I have overlooked. The remaining three are methodological objections.

1.5.1 What About Metaphysical Intuitions?

Thus far, we’ve considered empirical data given by the natural sciences and theoretical considerations as evidence for metaphysical theses. I have yet to say anything about intuitions, which are an evidential source distinct from empirical data and theoretical considerations. I take intuitions to be a priori seemings.

---

12For more on intuitions, see Huemer (2001).
These seemings can be perceptual (e.g., it seems that the sky is blue) or intellectual (e.g., it seems that 2+2=4), and provide prima facie justification for the truth of their content. A metaphysical intuition would thus be an intellectual seeming with metaphysical content. It is distinct from empirical data because it is a priori, and distinct from theoretical considerations since the intuition is not itself a theoretical virtue. The use of intuitions in philosophy is controversial, but, for the purposes of this objection, suppose that intuitions do provide at least prima facie justification. If we have metaphysical intuitions, then they provide some prima facie justification for believing some metaphysical theses. Perhaps when there is prima facie justification for a thesis, and that thesis can be defended from objections to it, then metaphysicians will be justified in believing that thesis.

This point is problematic for two reasons. First, it is highly doubtful that we have reliable intuitions about substantive metaphysical issues. Because metaphysics involves so many technical terms, many of the relevant concepts such as composition are not ordinary ones that people encounter. Even terms such as ‘world’ or ‘part’ that appear to have an established, ordinary sense are used in a subtly different sense by metaphysicians. Van Fraassen (1995) points out that because the terms have subtly different senses, metaphysical intuitions turn on privileging one sense over another. If van Fraassen is right, these senses are technical, and there is no reason to think that we have reliable intuitions about them at all. Indeed, it would be surprising if we did.

Second, on the view that metaphysics is continuous with the natural sciences, it is not uncommon for metaphysicians to hold that the next closest field to metaphysics is physics. According to this line of thought, physics is the most fundamental of the natural science, so if metaphysics is continuous with and even more fundamental than the natural sciences, the closest field to metaphysics is the most fundamental science. Suppose that this view is correct. People have incredibly
unreliable intuitions about physics. Part of why intuitions are so unreliable is that physics, as the most fundamental natural science, is removed from our ordinary experience and is comparatively abstract. It would be surprising if our intuitions about an even more abstract field were any more reliable. In fact, [Rose and Schaffer (2015)] tested folk intuitions about mereology, and found that folk intuitions are teleological. It turns out that folk intuitions generally exhibit a teleological bias, and that there’s a good debunking explanation for this bias. If Rose and Schaffer are right, there is empirical evidence that suggests that metaphysical intuitions are unreliable.

To be sure, there are worries about ethical intuitions as well. There is a historical association between teleology and what we ought to do that is encapsulated in Aristotle’s ethics and its philosophical descendants. However, these intuitions certainly are on no worse footing than our metaphysical intuitions. In many cases our ethical intuitions are better—at least insofar as there is a wider base of agreement. Even debunking arguments in metaethics are aimed at non-natural moral realism. Very few people are prepared to deny that there are at least some things that we ought to do or that some states of affairs are better than others. Furthermore, unlike the metaphysical case, many ethical concepts are ordinary ones. The obstacles affecting metaphysical intuitions do not apply to ethical ones. Again, my main proposal does not rely on ethical intuitions being the sole source of normative data. However, assuming that intuitions do matter evidentially, ethical ones are more reliable than metaphysical ones.

1.5.2 Why Not Other Data?

My positive proposal does not entail that empirical data, theoretical considerations, and normative data exhaust all of our evidential sources. Given that we’re interested in the subject matter of metaphysics, it might appear most natural
to search for metaphysical data. Above I argued that metaphysical intuitions are not reliable and do not provide evidence for metaphysical theses. Intuitions and theoretical considerations, including conceptual analysis, constitute all of the a priori evidence one could hope to have, so metaphysical disputes are not going to be resolved a priori. But perhaps there is data we can gain from experience, other than empirical data from the sciences, that bears on metaphysical disputes. There might be a special class of metaphysical data, or perhaps there is some entirely distinct class of data that bears on metaphysical disputes.

If there are other data available, my proposal would be potentially problematic. At best, the data would confirm what the normative data indicate. Overlooking the metaphysical data is non-ideal insofar as having more evidence is epistemically advantageous, but benign with respect to the truth goal. At worst, however, I might be overlooking a source that is more reliable than the normative one I’ve been discussing. In this scenario, failing to be exhaustive about evidential sources does subvert the truth goal. If an alternative source is more reliable than normative data, then there will be cases in which considering the additional evidence would result in a different verdict with respect to truth than considering the normative data alone would. We then ought to be using this other data to resolve metaphysical disputes.

My central claim—that normative data can help us make progress with respect to some metaphysical disputes—still stands even if it turns out that there is another source of evidence. An epistemically responsible person ought to consider readily available evidence, and normative data is readily available regardless of whether there are other types of evidence. Furthermore, it would be even better for metaphysics if there were these additional sources of evidence because they would improve the prospects of resolving metaphysical disputes. Unfortunately, it’s doubtful that there are additional sources of evidence.

First, it’s dubious that there is metaphysical data that we can gain from
experience. The subject matter of metaphysics is continuous with that of the natural sciences, since both are about the world. However, the natural sciences are empirical while metaphysics is not. If this is the distinction between the two, we shouldn’t expect that there are any observations we can make that will yield specifically metaphysical data.

Second, there might be an evidential source that is distinct from the sources we’ve considered thus far. One potential source of evidence is natural language. It may be that natural language falls under empirical data (I have no view on whether linguistics is relevantly similar to the natural sciences). But if we suppose that it is a distinct source, it may bear on metaphysical disputes. Perhaps ordinary language is metaphysically committing. Mackie (1993), for instance, argues contra van Inwagen (1990). While van Inwagen holds that ordinary language carries metaphysical neutrality, Mackie claims that it is committing. If Mackie is right, then when we say “There is a table,” in ordinary discourse, that provides at least prima facie justification for the metaphysical thesis that there is a table. As she puts it, when we observe and report that there are tables, “in the absence of further argument, we should take it that, in this respect, things are as they appear to be” (250).

However, there are at least two worries about this view. First, one might deny that ordinary language is metaphysically committing because a substantial amount of language is non-literal. Mackie herself offers three types of cases in which ordinary language is not literal. The aim of discussing these explanations is to show that the table case does not fit any of them. But even if that is so, the explanations themselves suggest that non-literal language is prevalent; it may very well be the case that for some metaphysical statements, the presumption should not be that they are literal. Second, one might concede that the aim of ordinary language is metaphysical commitment, but then deny that we are successful in this aim. This is especially
plausible given the way our ordinary language either is misleading or fails to track scientific truths. Given that metaphysics is continuous with the natural sciences, it would be surprising if ordinary language were better at tracking metaphysical truths, especially given that they are less observable than scientific ones.

In light of this discussion, it’s doubtful that natural language is a source of evidence that bears on metaphysical truths. But if there is linguistic evidence, or any other sort of evidence that bears on metaphysical theses, that would provide even more hope for achieving the truth goal. Normative data, along with this other evidence, would help us avoid metaphysical epistemicism. At worse, perhaps the normative data would conflict with the other evidence, and we would be no worse off with respect to the truth goal than we would be if we had only empirical data and theoretical considerations.

1.5.3 Isn’t Metaphysics More Fundamental than Ethics?

On some views, metaphysics is more fundamental than ethics. Some contemporary metaphysicians go so far as to define metaphysics as being the study of the fundamental. For instance, Schaffer (2009) holds that metaphysics is about what grounds what—which eventually bottoms out with a most fundamental ground. Similarly, Sider takes the project of metaphysics to be about the fundamental structure of reality. Suppose that it is true that metaphysics is about a part of reality that is more fundamental than ethics. Why should our judgment about fundamental rest on our judgment about what is less fundamental?

My proposal does not deny that metaphysics is more fundamental than ethics—indeed, it does not rely upon either being more fundamental than the other! Instead, my proposal relies upon there being a necessary connection between certain metaphysical facts and certain ethical facts. Those who hold that there are levels of fundamentality won’t want to deny this connection. For instance, Schaffer’s project...
is to explain how the more fundamental grounds the less fundamental. The grounding relation is such that if x grounds y, x’s existence is necessary and sufficient for y’s existence.

Furthermore, something’s being more fundamental metaphysically speaking does not entail that it is more fundamental epistemically speaking. For example, physical brain states might be more fundamental than their corresponding mental states, but my mental states are more accessible than my brain states (at least to me). And I can know that my brain states are changing in virtue of knowing that my mental states are changing. Even if ethics is less fundamental than metaphysics, it can still be epistemically permissible to use normative data to arrive at metaphysical claims.

1.5.4 Supplemental Principles

Earlier, I suggested that normative data rules out some theses in metaphysical debates. For instance, we might use the observation that there are scenarios in which agents are morally responsible for their actions to run a modus tollens argument against Hard Determinism:

1. If Hard Determinism is true, then it is not the case that there are scenarios in which agents are morally responsible for their actions.

2. There are scenarios in which agents are morally responsible for their actions.

3. Therefore, it is not the case that Hard Determinism is true.

(2) is our normative data, and (3) is our metaphysical progress. But what should we say about (1)? Presumably (1) is justified by a supplemental principle along the lines of:
Moral Responsibility: Moral responsibility cannot obtain without freedom.

But now we’re faced with a problem. Moral Responsibility appears to be a substantive metaethical claim, as opposed to a piece of normative data the way (2) is. Thus, what we have is an inconsistent triad between Moral Responsibility, (2), and Hard Determinism. Our normative data, (2), tells us that we must reject Moral Responsibility or Hard Determinism, but underdetermines which of the two we ought to reject. Furthermore, Moral Responsibility and Hard Determinism seem like good candidates for empirical and theoretical equivalence.

This objection is troubling, especially because it generalizes to other attempts to bridge the gap between normative data and metaphysical progress. Normative data by itself does not yield true metaphysical theses. Supplemental principles are required as well. These supplemental principles are epistemically troubling because the reasons for believing that there is normative data do not obviously apply to supplemental principle as well. For instance, suppose that we gain access to normative data by experiencing value. It’s then plausible that we can see that some scenarios, such as those in which people are morally responsible for their good actions, there is value. But it doesn’t seem that these observations provide evidence for Moral Responsibility.

There are a few ways to respond to this objection. First, it might be the case that true supplemental principles are knowable. Some, like Moral Responsibility, border on being conceptual truths. Plausibly, it is a precondition for moral responsibility that the act for which one is morally responsible is done freely. If that’s right, then sometimes supplemental principles such as Moral Responsibility are knowable a priori. A second possibility is that we can expand the reach of normative data to include these supplemental principles. Unfortunately, this response seems ad hoc unless there is a story about how we gain
normative data that includes supplemental principles. Perhaps if a strong enough
majority of ethicists agree upon a supplemental principle, it counts as a piece of
normative data.

Finally, it’s important to keep in mind that supplemental principles don’t have
to be completely unassailable. It’s helpful to compare this case to an analogous case
from the philosophy of science. Duhem (1914) observed that evidence provided by
empirical tests do not falsify hypotheses; instead, they falsify the conjunction of the
hypothesis, H, in question along with auxiliary assumptions, $A_1 \cdots A_n$. Thus, the
evidence is compatible with $\neg(H \land A_1 \land \cdots \land A_n)$, $\neg H \land A_1 \land \cdots \land A_n$, $H \land \neg A_1 \land \cdots \land A_n$, etc. Similarly in our case, perhaps the normative data doesn’t falsify
Hard determinism; instead, it falsifies the conjunction of Hard determinism
and Moral responsibility. But in science, we typically don’t think that there’s
massive underdetermination that results in scientific epistemicism (though some like
Duhem did take that to be what he had shown). This suggests that in science, it’s
epistemically permissible to rely on auxiliary hypotheses. A similar standard might
apply to our case. If so, we’re allowed to rely on supplemental principles like Moral
responsibility. The objection is as devastating to the project as Duhem’s point is
to the project of science.

1.5.5 Does Normative Data Presuppose Moral Realism?

Finally, one might worry that my appeal to normative data presupposes the truth
of moral realism. Moral realism comes in a variety of different flavors, but at its
core, it is a thesis on which moral and/or value properties exist. As such, it is
a metaphysical thesis, and a controversial one at that. I’ve claimed that belief in
controversial metaphysical theses cannot be justified by empirical data or theoretical
considerations, but that belief in some theses might be justified by normative data.
Thus, if it turns out that my appeal to normative data presupposes a controversial
metaphysical thesis, then the view I’m advancing is circular.

Fortunately, normative data does not presuppose moral realism. Instead, normative data is typically taken to be evidence for moral realism. If there is an epistemic connection between normative data and moral realism, moral realism is a (happy) consequence of normative data, rather than a presupposition of it. Take, for instance, the argument for moral realism from Oddie (2005). On Oddie’s account, desires are experiences of value. It then turns out that this data stemming from experiences of value in the world is best explained by a non-reductive moral realism. Huemer (2005) also gives an argument for moral realism based on normative data. For Huemer, we have moral intuitions rather than experiences of value. Still, these intuitions serve as normative data that counts in favor of moral realism. Even nihilists acknowledge that there is normative-like data, though they deny that it counts in favor of moral realism. For instance, Mackie (1977) acknowledges that many people take as evidence for realism that acting cruelly is wrong. He then offers an error theory for why this data does not entail that there are moral or value properties.  

The dialectic in the metaethics literature makes it clear that normative data does not presuppose moral realism. Instead, normative data is taken to be evidence that bears on the thesis that there are moral or value properties, even if it does not ultimately turn out that moral realism is true. Thus, the central claim of my paper—that normative data bears on some metaphysical theses—does not presuppose moral realism.

13 An intermediary position between moral realism and nihilism is moral relativism. If moral facts are relative to a culture, would metaphysical facts also be relative to a culture? This is an interesting question that I don’t have space to fully address here. However, there is a spectrum of responses that would be worth exploring in the future. For instance, one might think that only moral facts that have a wide base of agreement (and there are such facts across cultures) count as evidence for metaphysical theses. In this case, metaphysical facts would not have to be relativized to a culture. Alternatively, it may turn out that both moral and metaphysical facts are relative—though perhaps if one domain of facts is relative, it may not turn out to be surprising that the other is as well.
1.6 Ethics-First Metaphysics

I’ve argued above that normative data is an overlooked source of evidence that can help us resolve metaphysical disputes. In this section, I say more about what this amounts to with respect to making progress on metaphysical disputes. Following Dasgupta (2016), I refer to the general approach where we use the normative data to arrive at metaphysical conclusions as ethics-first metaphysics. To be clear, my claim is not that all of metaphysics should be done with an ethics-first methodology. Furthermore, the sense in which ethics comes first is purely epistemic. The methodology is most appropriate in cases where the relevant normative data occupies a more central position in our Quinean web of belief than the metaphysical thesis in question. This approach has limits, which I’ll discuss below. But it also has the virtue of providing a response to two of the challenges we started the paper with: that metaphysics doesn’t matter because disputes are irresolvable and because they have no bearing on our practical lives.

On one common way of approaching philosophical issues, metaphysical issues do not depend on ethical ones, either because metaphysics is more fundamental or because it is independent of ethics. When issues in one domain are connected to the other, ethics tends to defer to metaphysics. For instance, decision procedures that are designed to promote the self-interest of agents are neutral with respect to metaphysical questions such as what the correct view of personal identity is. Practical rationality takes for granted that there is identity, and then operates off of that starting point. Similarly, in the non-identity problem, the standard responses to the problem all take for granted that origin essentialism is true, despite the fact that that thesis is controversial amongst metaphysicians. Metaphysicians often embrace this general approach because it provides an easy way of explaining why metaphysics is so important. Metaphysics is needed to resolve issues, such as when personal identity holds, so that ethics can take place. Again, let’s borrow Dasgupta’s terminology and
refer to this approach as *metaphysics-first ethics*.

I’ve suggested that we take the reverse approach, and adopt an ethics-first metaphysics. Curiously, very few philosophers, including ethicists, have argued for this approach. The strongest stance against metaphysics-first ethics appears to be one on which ethics is independent of metaphysics. Parfit exemplifies this type of stance when he argues that we probably can’t know metaphysical facts (if there are any) such as the ones that bear on personal identity. Nevertheless, on Parfit’s view, we can focus on what matters and carry on with ethical pursuits regardless of the state of metaphysics.

Overlooking or resisting ethics-first metaphysics is a mistake. The few cases in which philosophers have taken this approach have resulted in metaphysical progress. The most clear case in which this approach is employed is the case of personal identity. Locke, for instance, used facts about blameworthiness and the principle that no one can be blameworthy for something another person does to develop his view of personal identity. More recently, Dasgupta argues that when we reflect on the non-identity problem, we should take seriously the ethical intuition that the person who conceives the less well-off child does something wrong. We should proceed by assuming that this intuition is reliable, and question instead the metaphysical claim—in this case origin essentialism. Doing so gives us a way to make progress in disputes about essentialism. [Schoenfield (2015)] also uses an ethics-first methodology in her recent argument for ontic indeterminacy (or the rejection of moral realism). She starts with a piece of normative data—namely, that there is moral indeterminacy. From this starting point, she argues for the claim that if moral realism is true, then there must also be ontic indeterminacy. Finally, there are hints of the ethics-first approach in the literature on composition. The Many-Brothers Determinism Argument presented by [Hudson (2001)] poses a challenge to responses to the problem of the many that concede that there are many, mostly-overlapping
The success of these cases suggests that the ethics-first metaphysics approach is worth pursuing in more cases. I explore one such potential case in the next section. Before getting to it, it’s important to note the limitations of this methodology. It’s doubtful that an ethics-first approach can resolve all metaphysical disputes. For instance, I doubt that this approach will resolve the dispute between nihilists and universalists when it comes to the composition of non-personal objects. Similarly, I doubt that the approach will answer the question of whether there can be coincident non-personal objects even though I do think it resolves the question of whether there can be coincident persons. In general, I suspect that the ethics-first approach will be applicable when the metaphysical dispute in question involves things that have to do with normativity. For instance, the approach is promising when the theses in question involve persons because persons are moral agents and patients. It also appears promising when it comes to reasons, since reasons carry normative force. This result makes sense given that normative data concerns value and what we ought to do.

1.7 Persons

Thus far, I’ve argued that an ethics-first methodology can help make progress on metaphysical disputes involving persons. In this section, I sketch how this might go in one particular kind of case. The section is intentionally merely a sketch—my primary purpose is to show that the ethics to metaphysics approach can do some important and interesting work. I focus on three related problems: the problem of coincident objects, the problem of plenitude, and what I call the personal problem of plenitude. I start by explaining the relationships between the three problems. As I understand them, the problem of plenitude is an extension of the problem of coincident objects, and the personal problem of plenitude is a special case of the problem of plenitude.
Next, I’ll turn to two questions that arise for all three problems. First, how many objects (or persons) are there? Second, if there is only one object (or person), which object (or person) exists? It may initially appear that since the three problems are structurally analogous, an answer to the two questions for one of the problems would generalize to the other two problems. However, contrary to initial appearances, the two questions can be answered with respect to the personal problem of plenitude, but not the other two problems. Reflecting on why the solution does not apply uniformly to all three problems reveals that there is something special about persons that makes ethics-first metaphysics a particularly appropriate approach to answering those questions.

1.7.1 Three Problems

Coincident objects present an interesting metaphysical puzzle. For instance, consider Lump, a piece of clay that can survive all changes to its shape, and David, a statue made from Lump that cannot survive drastic changes to its shape. Lump and David are perfectly coincident; they occupy precisely the same spatial location. They also share all of their categorical properties—they have the same mass, shape, color, etc. It is thus tempting to believe that there is only one object as opposed to two distinct ones. However, Lump and David differ with respect to their non-

\[\text{Gibbard (1975) presents the canonical statement of the problem as an argument for relative identity.}\]

\[\text{The story can also be told such that the statue and lump also occupy all of the same temporal locations. If the top and bottom halves of David are made separately then conjoined as the last step in statue making, David and Lump come into existence together when the two pieces are connected. They might go out of existence together as well if the statue shatters into pieces.}\]

\[\text{Use of the term ‘categorical’ is pervasive in the literature on material constitution and coincident objects. Unfortunately, the term also is notoriously unclear. Though there is no settled definition, there does seem to be a rough shared conception to which those who use the term adhere. For my purposes, the rough conception should be sufficient. I’ll take categorical properties to be non-modal, non-dispositional properties that include properties like shape and mass. In contrast, non-categorical properties are modal or dispositional ones, such as possibly having different shape or being fragile.}\]
categorical properties. Lump can survive being radically reshaped as a ball, but David cannot. Since Lump can survive radical reshaping and David cannot, it appears that via Leibniz’s law there are two things with two very different features—one that can survive reshaping and one that cannot.

This puzzle about coincident objects is one of the two paradigm disputes that Bennett holds are unresolvable. Like the debate over the composition of tables, this debate about coincident objects won’t be settled by empirical data. The natural sciences can tell us about objects’ categorical properties, but these facts won’t advance the dispute since we already know that David and Lump have the same categorical properties. The dispute also won’t be settled by appealing to theoretical considerations. One-thingers, those who hold that there is only one object, have the theoretical virtue of ontological parsimony. But multi-thingers, those who hold that there are coincident objects, have the virtue of ideological parsimony. The two positions are on a par when it comes to theoretical considerations. Thus, if epistemic data and theoretical considerations are the types of evidence that bear on this puzzle, the problem of coincident objects leads to metaphysical epistemicism.

The problem of plenitude is an extension of the problem of coincident objects. Consider a more detailed case for Lump and David being two distinct but coincident objects. Lump and David have different modal properties; Lump can survive radical reshaping but David cannot. Here’s another way to put this point. Both have the property of being statue-shaped, but they have this property in different ways. For David, this property is essential—David cannot survive lacking it. For Lump, this property is inessential, or accidental—Lump can survive losing it. Let a thing’s essence be the set of its essential properties. David and Lump do not have the same essence. On the standard conception of essence, objects are identical if and

17 For the purposes of this discussion, we can be neutral about whether a thing’s essential properties are the properties that it has necessarily or the properties that define what it is.
only if they have the same essence. Since David and Lump do not have the same essence, they are not identical. Thus, there must be two non-identical but coincident objects.

Once individuals are specified by their essences, we see that there can be far more than two coincident objects. Consider a more intolerant version of David, David’. While David can survive minor changes to its shape like having a deformed pinky finger, David’ cannot survive even that change. The essence of David’ includes essentially having a well-formed pinky finger, while the essence of David does not include that property. There might also be Lump’, the lump of clay that can survive not only being radically reshaped, but also being torn apart into smaller clay pieces. Specifying different essences results in a plenitude of coincident objects.

In order to stop this explosion of objects, we need what [Leslie (2011)] calls a principle of limited variety of essences. This principle would be a metaphysical one that rules out certain combinations of essential properties and thus cuts down on the number of essences. However, it is difficult to articulate such a principle in general, especially when it appears that properties are independent and can be recombined. To see why, consider an example from [Yablo (1987)]. Suppose that there is an object that has exactly five, modally independent essential properties. That object’s essence is constituted by those five properties. But why could there not be an essence that has only four of those properties? And why could there not be an essence for each of the four-property recombination?

I won’t delve into the literature that has formed around this issue. But if we step back and look at the debate, we can see that there are clear parallels between the debate about plenitude and the debate about coincident objects. In both debates, there is no empirical data that the natural sciences can offer that will settle the debate. Furthermore, the two positions are on a par with respect to theoretical virtue. Those who hold that there is some principle of limited variety of essences, like the one-
thingers, are virtuous with respect to ontological parsimony, but have more ideological commitments. Those who hold that there is a plenitude of objects, like those who hold that there are coincident objects, are not virtuous with respect to ontological parsimony, but do have a simpler ideology. Like the problem of coincident objects, debate over how to resolve the problem of plenitude is a paradigm metaphysical dispute that results in metaphysical epistemicism.

Finally, the personal problem of plenitude is a special case of the problem of plenitude. Just as David can endure more shape change than David’ can, there might be a person, Rebecca, who cannot survive ceasing to be a philosopher and another person, Rebecca’, who can. Rebecca and Rebecca’ share all of the same categorical properties, but differ with respect to one modal property. They coincide now and at all times they exist should it turn out that they never cease being a philosopher. Again, we can specify more profiles that are the same with respect to their categorical properties but differ with respect to their modal properties. Thus, just as we might puzzle over whether there is one or are many statues, we might puzzle over whether there is one or are many persons. This is the personal problem of plenitude. Like the other two problems, it appears that metaphysical epistemicism is the appropriate response to the question of how many persons there are—at least if empirical data and theoretical considerations are the only available evidence.

1.7.2 Two Questions

Two questions arise in each of the three problems. First, how many objects (or persons) are there? Second, if there is exactly one object (or person), which object (or person) is it? The first question is a paradigm case of a metaphysical dispute we should regard with metaphysical epistemicism. Knowing all the relevant scientific facts does not settle whether there is one object or multiple perfectly coincident objects. Furthermore, theoretical considerations do not settle the question either.
Theories positing only one object are on a par with respect to theoretical virtue to theories positing multiple objects. What the latter gives up in ontological parsimony, the former gives up in terms of ideological parsimony. The second question has not received as much attention as the first, but it seems fair to say that if scientific data and theoretical considerations do not provide us with sufficient evidence to resolve the first question, they certainly do not provide us with enough evidence to resolve the second.

I must confess that I do not hold out hope for resolving these problems by appealing to scientific data or a priori considerations for reasons given by Bennett that were discussed earlier in this paper. Nevertheless, I think that there is hope in at least the personal problem of plenitude if we consider normative data. I’ll start by looking at a response to the problem of the many considered by [Hudson (2001)]. This response inspires two arguments from normative data that answers the first question by concluding that there is exactly one person. I then propose that normative data also may give at least a partial answer to the second question. Which person exists is at least partially determined by facts about the agent’s practical identity.

1.7.2.1 How Many Persons?

In the three problems above, there are allegedly many perfectly coincident objects or persons. There is a related problem—the problem of the many—on which there are many overlapping, but not completely coincident, objects and persons. [Lewis (1993)] gives us the example of Tibbles the cat, who has 1000 hairs. If a hair, \( H_1 \), is removed from Tibbles, the resulting cat, \( C_1 \), is still a cat. *Mutatis mutandis* for the cats \( C_2 \ldots C_{1000} \) that result from removing hairs \( H_2 \ldots H_{1000} \). Including Tibbles, the problem of the many originates with [Unger (1980)], but I use Lewis’s example since cats are superior to clouds.
there are 1001 cat candidates. In fact, there are far more than 1001 cat candidates since there are cat candidates that result from moving various combinations of hairs. And, of course, we can remove more than hairs—e.g., tails, as Geach suggests—yet still have a cat. There are many, perhaps infinitely many, cat candidates that nearly perfectly overlap, and each of these candidates appears to be a cat. Using the same process for persons quickly yields the result that there are many, perhaps infinitely many, overlapping person candidates. Furthermore, each of these person candidates appears to be a person. The two questions that apply to our three problems apply to the problem of the many as well. How many objects, cats, or persons are there? If there is exactly one, which one of the candidates is it?

I want to focus on a promising answer to the first question, and develop it into an answer that applies to the personal problem of plenitude as well. I take the Many-Brothers Determinism Argument presented by Hudson (2001): 40 to contain the strongest case for there being exactly one person:

1. Suppose that I freely clap my hands at T.

2. If I freely clap my hands at T, then at least one of my M-brothers (say, Edward) freely claps his hands at T.

3. So, Edward freely claps his hands at T.

4. Necessarily, if Edward freely claps his hands at T, then I clap my hands at T.

5. If (i) A’s freely doing x at t entails B’s doing y at t, and (ii) A freely does x at t, and (iii) A is distinct from B, then B does not freely do y at t.

6. Hence, I do not freely clap my hands at T.

7. Reductio.
This argument is aimed at those who favor overpopulation solutions, or responses to the problem of the many that concede that there are in fact many people. It seems that each of the many is similarly situated with respect to freedom. But if (5) is true, the argument has the consequence that at most, one of the many is free. Perhaps none are free, but this contradicts evidence that we otherwise would have taken to be strong evidence for freedom.

My interest in this argument lies in the fact that (1) is (or is very close to being) a piece of normative datum. As I suggested earlier, there is normative data that provides strong evidence for the thesis that there are at least some instances of agents acting freely. Suppose that (1) is one such instance. That, combined with this argument, suggests that exactly one person acts freely. If Hudson is right about (5)—or something close to (5)—being true, then (2) must be false (since (3), (4), and (6) follow from the other premises). And the best explanation for (2) being false is that there is exactly one brother. If we take seriously that each of the many is similarly situated with respect to freedom, then there must be exactly one person.

Appealing to normative data provides the strongest case for there being exactly one person in problem of the many cases. If the Many-Brothers Determinism argument isn’t convincing, consider other less controversial pieces of moral data that we find in everyday scenarios. When I punch someone in the face, there is exactly one wrongdoer. It’s also the case that I commit one wrong, not many wrongs, one to each of many persons. When I donate to a less well-off stranger, I make one person better off. My act produces a finite improvement to the world, not an infinite one should it turn out that there are infinitely many overlapping people. People in monogamous relationships who make marriage vows are not mistakenly polygamous.

---

19 Hudson suggests that (5) is the only assailable premise, but that it is likely that it can be revised such that there are no counterexamples that arise in the case of non-overlapping persons. See pp. 41-2 for these other formulations.
This normative data provides the groundwork for two arguments, one deontic and one axiological. Here’s the deontic version:

8. Suppose that there is some action X such that I ought not to perform X at time t but I do perform X at t.

9. If I perform X at t, then one of my M-brothers (say, Edward) performs X at t.

10. So Edward performs X at t.

11. Necessarily, if Edward performs X at t, then I perform X at t.

12. If A’s performing X at t entails B’s performing X at t, and (ii) A performs X at t, and (iii) A is distinct from B, then it is not the case that B can not perform X at t.

13. Hence, it is not the case that I can not perform X at t.

14. If it is not the case that I can not perform X at t, then it is not the case that I ought not to perform X at t.

15. Reductio

(8) is our normative datum, and it’s more certain than the other premises. The only other premises are (9), (12), and (14). (12) appears analytically true. (14) follows from the widely accepted “ought implies can” principle, and may even count as normative data. (9), the analogue of (2) in the Many-Brothers Determinism argument, is thus the premise that is false. The best explanation for its falsity is that there is exactly one brother.

---

To be fair, there are those who reject that ought implies can. Perhaps they would take the argument to be a reductio against (14) and that principle. I have nothing to say in defense of the principle that has not already been said, and would direct those who reject the principle to the axiological argument, which does not rely upon it.
There’s also an axiological argument that yields the same conclusion:

16. Suppose for reductio that Edward has infinitely many brothers.

17. Torturing Edward is more disvaluable than punching Edward in the face.

18. If Edward has many brothers, then torturing Edward is infinitely disvaluable and punching Edward in the face is infinitely disvaluable.

19. If two acts are each infinitely disvaluable, then neither is more disvaluable than the other.

20. If torturing and punching Edward are each infinitely disvaluable, than neither is more disvaluable than the other.


In this argument, (17) is our normative datum. (19) is the only premise that is independent of the supposition that there are many rather than one, and it’s likely analytically true. Again, turning to normative data leads us to the conclusion that the correct response to the problem of the many, at least when it comes to person, is to conclude that there is exactly one person.

These arguments also apply to the personal problem of plenitude. All we have to do is let the brothers include coincident persons as well as nearly completely overlapping ones. The best evidence for there being exactly person as opposed to multiple coincident persons is that normative facts entail that there is exactly one agent. To be sure, one who holds this position faces the task of answering the further question of which person is the agent. But that there is this further difficult question that arises from concluding that there is exactly one person should not dissuade us from taking the normative data seriously. In any case, should it turn out that there
is a way of answering the further question of which person is the person (which is what I aim to do in the next section), this worry is unfounded.

Finally, it’s worth mentioning that the arguments might also imply that there is exactly one cat in the Tibbles example. Plausibly, cats are moral patients, and my act of petting a cat and making it happy does not involve infinitely many good acts, one corresponding to each cat I’m petting. It seems doubtful that similar reasoning applies in cases of objects that are neither moral agents or patients (though if it does, then that would show that the ethics-first approach is even more significant than I’ve proposed). The limits of the ethics-first approach in these cases are tied to moral subjects, which makes sense given the methodology I’ve proposed.

1.7.2.2 Which Person?

If I’m right thus far, then normative data gives us a resolution to the first question as it applies to the personal problem of plenitude. But now that the first question has been settled, the second naturally arises: which of the candidate persons is the person? Recall two coincident person candidates, Rebecca and Rebecca’. The former cannot survive ceasing to be a philosopher, but the latter can. If there is exactly one person, is that person Rebecca or Rebecca’? There are a two general ways one can go on this question. I’ll lay out the range of options, and then offer my own favored solution, which incorporates normative data.

First, one might hold that in cases of plenitude, each of the coincident objects is an equally good candidate for being the object. In our example, both Rebecca and Rebecca’ have equally good claim to being the person. Since both have equally good claim, it would be arbitrary for one, rather than the other, to be the person. Thus, the answer to which person is the person must not privilege one candidate over the other.

There are a few different forms a non-arbitrary answer might take. First, one
might take a Lewisean-inspired approach and hold that both are the person, but since they are coincident in many worlds, the are almost one—and being almost one is close enough to our answer to the first question regarding how many persons there are. Second, one might hold that since the two are coincident, it’s indeterminate which of the two exists. Finally, one might hold that a disjunctive, person-like entity exists. Dasgupta (2016) holds such a view. On it, there is a plenitude of candidate persons (e.g., Rebecca and Rebecca’ both exist), as well as a person-like entities, such as the disjunctive entity that exists as long as Rebecca or Rebecca’ exist. On Dasgupta’s view, our practical concerns dictate which of the candidate entities exist, and sometimes, our practical concerns dictate that the disjunctive candidate exists. So given that there is exactly one person, and that the coincident person-candidates have equally good claim to being the person, it may be that the person is a disjunctive, person-like entity.

Second, one might hold that in cases of plenitude, not all of the coincident objects are equally good candidates for being the object. For instance, it might be the case that Rebecca and Rebecca’ do not have equally good claims to being the person. One who holds this view must explain why one candidate has the better claim than the other. It may seem like the nature of the problem of plenitude is such that the candidates do in fact have equal claim. After all, those who wish to avoid plenitude attempt to find some grounds on which some candidates fail to be as good as others. The less good candidates are disqualified, and even if not enough candidates are disqualified to get to one object, the plenitude is significantly mitigated. There’s something right about this point, and I will go on to argue that some candidates are better than others and that this eliminates at least some of the plenitude.

It’s worth noting a couple of things about this second general way of answering which person is the person. First, because we started by using normative data to determine that there is exactly one person, the question of whether some candidates
are better than others has not yet arisen, so it makes sense to now consider whether some candidates can be eliminated from contention as the candidate. Second, the general approach to the problem of plenitude is to provide some metaphysical grounds that explain why the world is not as plenitudinous as the problem would have us believe. But like the problem of coincident objects, the competing metaphysical theses are subject to metaphysical epistemicism. Empirical data does not settle the dispute, and neither do theoretical considerations. Theories positing fewer objects have the virtue of ontological parsimony, but those that allow for plenitude have the virtue of ideological parsimony. In short, there is something right about approaching the question of which person exists by thinking about what might make one candidate better than another. However, it is dubious that metaphysical considerations can in fact provide justification for believing that one candidate is better than another.

My proposal is that normative data can help us with respect to the question of which candidate person is the person. Facts about what we ought to do include facts about what a person ought to do from the standpoint of self-interest. Though this might come apart from what a person ought to do all-things-considered, what they ought to do from the standpoint of self-interest is still a piece of normative data. Start by supposing Rebecca and Rebecca’ are equally good candidates for being the person. Now suppose that the person is faced with a decision where one outcome involves remaining a philosopher and the other involves never being a philosopher again. Perhaps the person is deciding whether to become a boxer, and though that future would be enjoyable, the brain damage it would involve would preclude doing philosophy. Rebecca is such that she cannot survive ceasing to be a philosopher, so for her, opting for the boxing future is tantamount to opting for a future that does not involve her existence. Assuming that the future involving philosophy is a positive one, Rebecca ought to opt for that one from the standpoint of self-interest. On the other hand, Rebecca’ can survive ceasing to be a philosopher, so for her, she ought to
opt for whichever future involves greater value since she survives in either outcome. Suppose that the boxing outcome involves more value for the person than the non-boxing outcome. From the standpoint of self-interest, Rebecca’ ought to opt for the boxing outcome. Thus, what the person ought to do, then, depends on whether the person is Rebecca or Rebecca’.

If there is a determinate fact regarding what the person ought to do, then there is a normative fact that disqualifies either Rebecca or Rebecca’ from being the person. Developing a full account about these types of facts is beyond the scope of this paper, but it is worth noting that there are some accounts that attempt to capture these normative facts about persons. For instance, Schechtman (1996) distinguishes between two questions about persons, the ‘reidentification’ question and the ‘characterization’ question. The former asks the standard personal identity question: what is it for a person at \( t_2 \) to be the same person at \( t_1 \). The latter asks what characteristics are “truly those of the person” (73). These are the characteristics that make and express who a person is. Being a philosopher might be one of these central characteristics for a particular person, or it might not be. But if it is, then there is a normative fact such that preserving philosophical ability is in the person’s self-interest.\(^{21}\) To be sure, normative facts may not get us all the way to one best candidate. However, it can help us make significant progress. Furthermore, where there are remaining candidates that have equally good claim to being the person, there will be no practical difference between these candidates. Because there is no practical difference, having a disjunctive person-like entity or having it be indeterminate which of the remaining candidates is the person is an easier consequence to accept. Thus, normative facts can help answer the second

\(^{21}\)Schechtman (1996) endorses a narrative self-constitution view on which personal identity is created by a self-conception that is narrative in form. I’m not committed to her view, but I do think it provides a good illustration of what these normative facts might be and how they would not be determined by metaphysical considerations.
question, as well as the first.

I’ve argued in this section that normative data helps resolve the problem of personal plenitude. But the ethics-first metaphysics approach does not generalize to the problems of plenitude and coincident objects. Persons (and perhaps other sentient creatures that qualify as moral patients) have normative profiles. Evaluatively, their lives can go better or worse, and deontically, there are acts that they ought or ought not perform. The same is not true of clouds, tables, and other ordinary objects. The metaphysics of persons has an extra evidential source that the metaphysics of ordinary objects does not. Metaphysicians ought to take advantage of this source and employ the ethics-first approach when it comes to person.

1.8 Conclusion

In this paper, I’ve argued for an ethics-first metaphysics methodology. It won’t resolve all metaphysical disputes—we might still be stuck with metaphysical epistemicism when it comes to most disputes about composition and coincident objects. But we can make progress with respect to some of them, such as disputes about persons. Disputes where normative data is relevant should be reexamined in light of normative data. Doing so allows us to make progress with respect to some metaphysical disputes and answers the challenge posed by metaphysical epistemicism.

Furthermore, the ethics-first approach shows that metaphysics is worth doing because it is connected to our practical lives. It’s more intimately intertwined with ethics than is typically realized. Admittedly, the approach I’ve argued for makes metaphysics appear to be the handmaiden of ethics. However, there will be cases in which the metaphysics-first approach is appropriate. For instance, metaphysics speaks to possibility and what we can do constrains what we ought to do. The point
to take away from this paper is that neither metaphysics nor ethics should be done in isolation. The intersection of the two contains the resources to make significant philosophical progress in both fields.

1.9 Cited Works


Abstract. In “Essence and Modality,” Kit Fine presents counterexamples to the modal account of essence that are widely regarded as successful. As a result, the modal account has fallen out of favor, and the trend in metaphysics is to move towards a definitional account of essence on which only some of a thing’s necessary properties—those that “define what it is”—qualify as essential properties. This paper argues that abandoning the modal account for the definitional one is too hasty. The modal account can be repaired by adding two modifications based on determinacy and sparseness. These modifications result in a revised modal account that avoids Fine’s counterexamples. Furthermore, unlike other attempts in the literature to modify the modal account, these modifications can be analyzed in purely modal terms. The revised modal account proposed in this paper is thus fully reductive, preserving the primary virtue of the earlier modal account.

2.1 Introduction

Socrates has many properties. Some of these are necessary, or such that he cannot lack them in any possible scenario, while others are contingent, or such that he can lack them. For example, Socrates has the properties of being short, snub-nosed, and human. While being human is a necessary property for Socrates because Socrates must be human in all possible scenarios, being short and snub-nosed are contingent properties. Had Socrates had the technology of the 21st century available to him,

\[1\] For the purposes of this paper, I take for granted that the property of being human being is a necessary and essential property. If, like Plantinga, the reader is inclined to think that Socrates could have been an alligator (or, more accurately, could have had an alligator body) and that the property of being human is not essential or necessary, the reader can substitute his or her favored prototypical
he could have undergone hormone treatment as a child or had plastic surgery, and thus had a different stature and nose.

In addition to necessary and contingent properties, Socrates also has essential and inessential properties. For preliminary purposes, let’s very roughly conceive of a thing’s essence as “what it is” to be that thing. So, for instance, the essence of Socrates provides an answer to the question “what is Socrates?” or “what is it to be Socrates?” (Since the purpose of this paper is to arrive at a precise, correct account of essence, we’ll begin with the intentionally rough conception to start the discussion without begging the question in favor of a particular conception.) Essential properties are those that make up Socrates’s essence, while inessential, or non-essential, properties are ones that do not make up Socrates’s essence. The properties of being short and of being snub-nosed are undoubtedly inessential properties as well as contingent ones. They are properties Socrates could lack, and thus do not get at what it is to be Socrates. The same is true for all other contingent properties. In light of the fact that all contingent properties are inessential, are all inessential properties contingent? Or to put it another way, are all necessary properties essential properties?

According to philosophers who favor a modal conception of essence, the answer to this question is ‘yes.’ On the other hand, philosophers who favor a definitional conception of essence answer the question negatively. The modal account of essence rose to prominence in the decades following Quine’s anti-essentialism because it offered a simple, fully reductive story about what a thing is. But since Fine’s essential property for that one. Also for the purposes of this paper, I set aside anti-essentialism and assume that things, such as Socrates, have essential properties.

For Correia (2006), the two questions are different. The former gets at an “objectual essence,” or what that particular object is, while the latter is a “generic essence” and gets at what it is to be a certain sort of thing. For our purposes, we need not make the distinction. The answers to “what is Socrates?” and “what is it to be Socrates?” will both be given by referencing his essence. The answer to the former will be “a thing that has such-and-such essence” and the answer to the latter will be “to have such-and-such essence.”
“Essence and Modality,” the modal account has fallen into disrepute and is no longer regarded as a viable account of essence. On the current mainstream view, the definitional conception is superior because it can do important metaphysical work that no extension of the modal conception can.

My goal in this paper is to revive the modal account of essence and restore it to its former glory. Though Fine and the defenders of the definitional accounts are right to point out flaws in the traditional modal accounts of the 1960s and 70s, it would be hasty to completely abandon the modal conception. With two modifications, we can have a revised modal account that is extensionally similar to a definitional account. Such an account would be significant because it would avoid Fine-style counterexamples without taking on some of the drawbacks of definitional accounts. If the modifications to the revised account could be given in purely modal terms—which is what I aim to do—it turns out that the revised modal account possesses theoretical advantages that the definitional one does not. This last point would be a happy result for the modal conception of essence since it would suggest that there is a modal account that is not only on a par with, but superior to, definitional accounts.

Before offering my positive view, I start by getting clear on what the competing conceptions of essence are (§2.2) and why the definitional conception is thought to be superior (§2.3). Next, I make some brief methodological comments (§2.4). I then offer my positive proposal (§2.5), on which all of a thing’s necessary properties still count as essential ones. The revision to the standard modal account will come in two stages: determinateness and sparseness. On my analysis of determinateness, a property is more (or less) determinate in proportion to how narrow (or wide) the scope of what a thing that has that property could be. A thing’s essential properties still are all and only its necessary properties, but these essential properties come in degrees corresponding to determinateness. On my analysis of sparseness, necessarily coextensive properties are the same property. This analysis of sparseness

53
is more serviceable than other competing analyses of sparseness in the literature on essence because it is given in purely modal terms. Given these two modifications, the extension of a thing’s most essential properties by the lights of the revised account will be the extension of a thing’s properties given as essential by the definitional account. Finally, since my revised modal account is given in purely modal terms, it is fully reductive and offers two theoretical upshots, parsimony and non-arbitrariness (§ 2.6).

2.2 Two Conceptions of Essence

On the simplest modal conception of essence, essentiality is analyzed in terms of modality. For example, Plantinga (1974) offers this sort of account when he says that “an object x has property F essentially if and only if there is no world in which x has not-F” (60). This conception of essence is appealing because necessary properties are ones that Socrates must have, and this necessity appears particularly salient. Furthermore, many necessary properties, such as the property of being human, get at what Socrates is. For ease of reference, let us formulate the simple modal conception of essence (‘SMC’) as follows:

\[(SMC) \text{ Where } F \text{ is a property, } F \text{ is essential to } a \text{ iff necessarily, } Fa \text{ if } a \text{ exists.}\]

Socrates’s essential properties are his necessary properties, and vice versa, since his necessary properties are the ones that he will have so long as he exists.

The most significant strength of SMC is that it offers a fully reductive account in virtue of analyzing essence in terms of modality. Reduction is desirable for at least two

---

\[^3\text{I've given the existential formulation. There are also other ways of formulating the modal conception (e.g., identity), but since all are vulnerable in the same way to the definitional challenge, having just the existential formulation as our exemplar of the simple modal conception serves our purposes.}\]
reasons. First, as Cowling (2013) notes, reduction is more ideologically parsimonious than adopting new primitives. Second, as Denby (2014) argues, essence is exactly the sort of thing we would expect to be reducible. Essence doesn’t seem to be primitive, and if it were, it would not be clear enough to provide a solid grasp of related notions. But despite its elegant, reductive analysis of essence, many have argued that SMC is inadequate. It is now widely believed that a conception of essence modeled on definition is superior to the modal conception.

On the competing definitional conception, though all essential properties are necessary, not all necessary properties are essential. As its name suggests, the definitional conception of essence likens the essence of a thing to its definition; just as a definition concisely and informatively explains what a word means, a thing’s essence explains what that thing is in a concise, informative way. As Fine (1994) puts it, “just as we may define a word, or say what it means, so we may define an object or say what it is” (2). Crucially, defining and explaining what a thing is does not involve listing all of that thing’s necessary properties. Let us formulate the definitional conception as follows:

\[(\text{DC}) \text{ F is essential to a if and only if F is part of the “real definition” of a.}\]

Unlike SMC, DC does not count all of a thing’s necessary properties as essential since not all of them are part of the thing’s definition. Also unlike SMC, DC does not have a fully reductive analysis since ‘is part of the real definition’ is taken as primitive. Even if one attempts to analyze ‘is part of the real definition’ in terms of fundamentality or metaphysical priority, that analysis will require taking fundamentality as a primitive.

However, definition is more fine-grained than modality. This fine-grainedness allows for a narrower conception of essence that is thought to provide a better account of what a thing is. Returning to the case of Socrates, many of his necessary properties
do not quite get at what he is. For instance, Socrates is not identical to his student Plato, and never could be. Thus, he has the necessary property of being distinct from Plato. But this property is not particularly informative with respect to what Socrates is; many other diverse things such as Aristotle and Mt. Olympus have this property as well. There are even more necessary properties that seemingly have nothing to do with what he is such as the property of being the sole member of Socrates’s singleton set and the property of being a man or a mountain. This problem is exacerbated by properties relating to necessary truths. Socrates also has the property of being such that 2=2. Since 2=2 in all possible worlds, and thus in all worlds in which Socrates exists, Socrates is always such that 2=2. In fact, any other necessary truth can be substituted in for “2=2” to yield yet another property he has necessarily. In light of cases like these, there seem to be necessary but inessential properties. Consequently, there is a case for moving towards a definitional account of essence.

2.3 The (Alleged) Virtues of Definition

The model of definition is more narrow than that of necessity. While SMC counts all of a thing’s necessary properties as essential, DC excludes some necessary properties from a thing’s essence. This exclusivity purportedly makes DC superior. This section highlights the alleged virtues of the definitional model.

2.3.1 Seemingly Inessential but Necessary Properties

Fine (1994) offers a wide range of cases in which a thing’s necessary properties do not appear to be essential. I’ll break these cases up into two broad groups. I employ these groups out of convenience—as we’ll see, they line up with the stages of my positive proposal—and not because I take the distinction to be metaphysically significant.

The first group of necessary but seemingly inessential properties is that of
non-discriminatory properties. Recall that Socrates necessarily has the property of
being such that 2=2. He also necessarily has the properties of being red or not red,
being self-identical, and being such that 1+1=2. All of these properties are ones
that all things have necessarily. They don’t get at what Socrates is because they
don’t distinguish him from any other objects. Disjunctive and negative properties
can also be non-discriminatory. A paradigm case of a non-discriminatory
disjunctive property is the necessary property of being P or ¬P. In the case of
disjunctive and negative properties, non-discrimination can come in degrees. There
are marginally more discriminating properties such as being a man or a mountain,
being non-mountainous, or being not-Plato. All of these are still too broad to give a
good sense of what Socrates is. More generally, non-discriminatory properties are
problematic in this context because they are not specific enough to get at what a
thing is.

A special case of non-discriminatory necessary properties, the property of
existence, is worth mentioning. I keep existence separate from the other cases
because it is peculiar and potentially controversial given one’s prior commitments
regarding existence. For instance, a Meinongian might hold that there are some
things that do not exist, and Kantian might hold that existence is not a property.
But if we set aside those sorts of views, then according to Fine (1994), counting
existence as an essential property is problematic because trivially, Socrates exists in
every world in which Socrates exists. He necessarily has the property of existing.
There is something peculiar about the property of existing being essential to
Socrates, but modal accounts of essence are committed to existence as an essential
property if all necessary properties are to be essential ones. Similarly, it turns out
that all things have the property of existence essentially since they all have that
property in all scenarios in which they exist. Thus, existence is also
non-discriminating and faces the worries that the other non-discriminatory
properties do.

The second group of seemingly inessential necessary properties is that of necessarily coextensive and derivative properties. Necessarily coextensive properties are properties that are necessarily coinstantiated: properties P and Q are necessarily coextensive if and only if necessarily, P is instantiated when and only when Q is. For example, trilaterality and triangularity are coextensive since they are both instantiated in all and only triangles. Derivative properties are coextensive properties that exhibit a dependence relation where one property is more fundamental and the other is derived from it. (All derivative properties, as I define them, are coextensive, but I leave open the possibility that the converse does not hold.)

Unlike the properties in the first group, coextensive and derivative properties can be perfectly discriminating. Socrates, for instance, has many necessary properties that are unique to him, such as the property of being Socrates and being the sole member of Socrates’s singleton set. But all of these properties are coextensive, and some of them, purportedly, are derivative. Take being Socrates and being the sole member of the singleton set containing Socrates. Necessarily, the singleton set containing Socrates exists if and only if Socrates exists. Thus, an essential property of the singleton is the property of having Socrates as its sole member. This result is intuitive. However, because Socrates only exists when the singleton exists, the modal account also yields the result that the property of being the sole member of Socrates’s singleton set is essential to Socrates. And according to champions of definitional accounts, this result is peculiar because Socrates is metaphysically prior to his singleton. As Fine (1994) puts it, “it is no part of the essence of Socrates to belong to the singleton” (6). The problem can be iterated with the singleton containing Socrates’s singleton, the singleton containing the singleton containing Socrates’s singleton, and so forth. On the definitional
2.3.2 Virtues of Exclusivity

What are the virtues of exclusivity when it comes to qualifying as an essential property? First, there is something worrying about counting the non-discriminating and coextensive properties as essential. Imagine Pat who is taking a philosophy class for the first time and who asks the professor what it is to be Socrates. Now suppose that Pat’s professor answers by saying that it is to be red or not red, be self-identical, and be such that 1+1=2. Though Pat’s professor has said some true things about what Socrates must be, none of these properties really get at what Socrates is because they don’t distinguish Socrates from other objects. Pat would be justified in feeling as if his question has not been answered. The same rationale holds for the slightly more discriminating properties such as being distinct from Plato or being a man or a mountain. Though these properties are slightly more discriminating than the necessary ones, they still are not discriminating enough to be sufficiently informative.

Now imagine that Pat’s professor gives a more discriminating answer and says that to be Socrates is to be the sole member of his singleton set. Again, if Pat were to feel that his question has not been answered, his feelings would be apt. Though that property is necessary, it is a round-about way of getting at what it is to be Socrates. The more narrow definitional conception of essence fits our intuitions about which properties answer the question of what it is to be Socrates.

Second, because it is more narrow, the definitional conception of essence is a conception on which essence does important metaphysical work. For Fine, one of the upshots of the definitional conception is that essence is able to give an account of ontological dependence. For instance, with the definitional conception of essence, we can have an account where “one object depends upon another...if its essence
prevents it from existing without the other object” (Fine 1994 2). This conception of essence might also help with other issues connected to metaphysical explanation, such as substance or grounding. These advantages (allegedly) flow from the narrower definitional conception and cannot be had by the broader modal conception. Recall the case of Socrates and the sole member of his singleton set. Plausibly, Socrates is more fundamental than this singleton. Perhaps his singleton even depends upon him. On modal conceptions, being the sole member of his singleton set is essential to Socrates and being Socrates is essential to the sole member of Socrates’s singleton set; there won’t be a asymmetry between Socrates and his singleton that captures the difference in fundamentality. But on the definitional conception, being the sole member of his singleton is not essential to Socrates, so there is an asymmetry in essences that yields a difference in metaphysical priority.

Third, essence is a useful tool for individuating amongst kinds and, potentially, individuals. In the case of kinds, the essence of a kind explains what it is for a thing to be of that kind. For instance, on Aristotle’s definition, humans beings are rational animals. The combination of rationality and animality explain what it is to be human. Furthermore, the differentia of rationality distinguishes human animals from other animals. If we suppose that the properties of being rational and being an animal are the essential properties of what it is to be human, we can see that the essential properties not only tell us what humans are but also how they are different from other animals and other kinds of things. The case of individuals is more controversial, but if there are individual essences, then they would provide an answer for what it is to be a particular individual. Just as essence could specify what type of thing something is, it could also specify what particular it is.

---

4 Whether essences can individuate individuals is controversial. Koslicki (forthcoming), for example, argues that unless essences include haecceities or forms, they will not be able to individuate. I won’t take a stance on this issue in this paper since if there are individual essences, they’ll function similarly to kind essences in a way that does not change the character of the dispute between the modal and definitional conceptions.
Finally, one might believe that part of the project of metaphysics is to “carve the world at its joints.” On the assumption that there are in fact joints, one theory is better than another if it does a better job at joint carving. For instance, a theory formulated in terms of blue and green is better than a theory formulated in terms of grue and bleen if blue and green are joint-carving properties and grue and bleen are not. Interestingly, both theories still might be strictly speaking true, but the theory given in terms of blue and green is superior because it is more apt. Similarly, we might think that the narrower conception of essence is more apt—perhaps because it represents ontological dependence relations, which are part of the fundamental structure of reality. If this is right, then the definitional account of essence is superior to the modal one if it lines up better with reality’s joints. Essence on the simple modal account is too broad and fails to match these joints. Taken with the previous points, these four points—exclusion of necessary but seemingly inessential properties, ability to do metaphysical work, individuation, and joint carving—present some compelling reasons to favor a definitional conception of essence over a modal one.

2.4 A Methodological Aside

Before diving into the heart of this paper, there are some methodological preliminaries that need clarification. For starters, it is important to see that the dispute between these two conceptions of essence is not merely verbal. One might worry that ‘essence’ is a technical term that has been introduced by philosophers. If this is right, then what essence ends up being is a matter of stipulation, and there is no real dispute between the two conceptions. They are merely philosophical tools that have been introduced for the purposes of the projects with which they are associated.

5 Joint-carving’ can be traced back to Plato’s Phaedrus. For a more contemporary discussion, see Sider (2011).
There are at least two reasons to reject this diagnosis of the dispute. First, the parties to the dispute do not take themselves to be engaged in a merely verbal dispute. For instance, both Plantinga and Fine are clearly responding to Quinean anti-essentialism. Quine (1960), who was skeptical of *de re* modality, rejected essentialism altogether. Part of Plantinga’s project was to respond on the behalf of *de re* modality and thus modal essence. Fine gives a different response on which he wants to preserve essence because of their special role in metaphysics. Fine, as we’ve seen in §2.3, is also responding to the modal conception.

Furthermore, new parties to the dispute also take it to be substantive. Cowling (2013), for instance, offers an interesting diagnosis of the three-way dispute between the anti-essentialist, definitional conception of essence, and the modal conception of essence. Others like Gorman (2005) and Wildman (2013) propose revisions to the modal account because of the challenges offered by Fine. They also take themselves to be in a meaningful dispute with Fine.

Second, all parties agree that what is at issue is whether essences have a special role—namely, the answering of ‘what’ questions—to play in metaphysics. The anti-essentialist, like Quine, will deny this, mostly because he wants to reject *de re* modality and holds that something else, like nature, can answer these questions. Those on the other side who hold that there are essences will say that essence does provide an answer to metaphysical ‘what’ questions, and then fight amongst themselves about whether the modal or definitional conception provides the better answer. So, if the defender of the definitional conception is correct about the benefits discussed in the previous section, that will be reason to think that the definitional conception is better than the modal one.

Now that we have a clearer view of what the dispute is about, we can see how

---

6See Cowling (2013), especially section 5, for an illuminating diagnosis of the Quinean, Finean, and modal essentialist positions.
the champion of the modal account might respond to the alleged advantage of the definitional conception. First, she might counter each of the points in favor of the definitional conception individually. This approach has been taken and has some merit. When it comes to the necessary but seemingly inessential properties, the champion of the modal account might claim that some of them, such as the properties everything has necessarily, are not really problematic. Della Rocca (1996), for instance, argues that properties that all things necessarily have, and all properties that follow from the ones that all things necessarily have, simply are not candidates for essentiality. Perhaps the necessary property of existence only seems weird because it is inadvertently conflated with necessary existence. The champion of the modal conception might also point out that modal essences are equally good at individuating. Or she might argue that grounding, fundamentality, joint carving, and related notions are unintelligible, and that because they are unintelligible, it is not a virtue of a conception of essence to fit those notions.

These responses have merit, but there is another way the champion of the modal conception might respond. She might revise the modal account such that the revised modal account is extensionally similar to the definitional accounts. An extensionally similar account would also draw distinctions amongst the necessary properties of a

---

7Chisholm (1967) suggests that “x has E [essential properties] in every possible world in which x exists,” and “for every y, if y has E in any possible world, then y is identical with x” (6). On this picture, the instantiation of essential properties in a thing turns out to be sufficient for that particular thing’s identity. Plantinga (1974) asserts a similar point: “the essence of Socrates is a property (or group of properties) that Socrates has essentially and that is unique to him. We could say that an essence of Socrates individuates him; it must be a property nothing else has” (60). If Socrates’s essence, or all of his essential properties, is instantiated in a world, then Socrates exists in that world since it is impossible for any other thing to have his essence. As noted earlier, whether there are individual essences is controversial, but if there are, it seems that modal accounts are well-equipped to pick out individuals.

8Though there is no shortage of philosophers who defend notions like grounding and fundamentality (e.g., Schaffer (2009); Sider (2011)), there are plenty who think that it is either unintelligible (e.g., Hofweber (2009)) or too board a notion to do serious metaphysical work (e.g., Koslicki (2015); Wilson (2014)). If one is unfriendly to grounding and fundamentality—especially if one thinks that grounding is unintelligible—then being joint-carving, for instance, will not be a virtue since it is not even intelligible.
thing and thereby provide a narrower conception of essence. Indeed, there have been recent attempts to revise the modal account in this way. This is the route I will take as well.

2.5 The Revised Modal Account

I now turn to my proposal for how the modal account ought to be revised. In a nutshell, essential properties can be sorted according to determinateness and sparseness. This proposal departs from competing revisions to the modal accounts in two significant ways. First, via determinateness, essentiality turns out to be scalar rather than binary. All of a thing’s necessary properties are its essential properties, but some are more essential than others. Second, I offer an account of sparse properties. But unlike other sparse accounts that attempt to cut down on the candidate essential qualities by disqualifying certain types of properties from being essential ones, my account cuts down on the candidate essential properties by holding that there are far fewer properties than initially thought. My account adopts a view of sparseness on which all necessarily coextensive properties are the same property. Since the analyses of determinateness and sparseness are purely modal, my revised account is also purely modal and retains the virtue of reduction that the simple modal account has. In what follows in this section, I elaborate on these two parts of my proposal and explain why the way in which they depart from other revisions to the modal accounts are advantageous.

2.5.1 Stage 1: Determinateness

In the first stage of my proposal, necessary properties are ordered according to determinacy. A property is more (or less) determinate in proportion to how narrow

---

9 E.g., Denby (2014); Gorman (2005); Wildman (2013).
(or wide) the scope of what a thing that has that property could be. At the most
narrow extreme, a property is maximally determinate if there is only one kind (or
individual) that a thing could be given that it has the property in question since the
scope of what that thing could be includes just one kind (or individual). For instance,
the property of being Socrates is maximally determinate since given that a thing has
that property, the scope of what that thing could be—namely, Socrates—is extremely
narrow. The property of being Socrates or Plato is slightly less determinate since
given that a thing has that property, it the scope of what it could be—Socrates or
Plato—is slightly wider. The property of being human or mountain is even wider,
and thus even less determinate. At the other extreme, properties that all things have,
like being such that 2=2, are the least determinate properties since given that a thing
has that property, the scope of what that thing could be covers all things. Like the
traditional modal account, this revised account counts all necessary properties as
essential ones, and vice versa. However, unlike the traditional account, which only
lists the essential properties, the revised account orders the essential properties in
degrees in accordance with determinacy. The more determinate a property is, the
more essential it is.

Significantly, determinacy is not merely an accidental feature of a world. Imagine
a world where the only things that exist are humans. In this world, all things have
the property of being human as well as the property of being such that 2=2. Because
the properties are instantiated in the same number of things, one might be tempted
to think that they are equally determinate in this world. However, the property of
being human is still more unique because a particular thing’s having that property
limits the scope of what that thing could be to just humans. The property of being
such that 2=2 does not limit the scope of what that thing could be whatsoever.

Thinking about determinacy in terms of possibility leads to a potential worry for
its analysis. To explain this worry, I’ll use possible worlds heuristics, though I’m not
wedded to any particular conception of possible worlds and the force of the worry and my response do not turn on it. Given sufficiently many worlds, there may be an infinity of humans and an infinity of animals. The scope of being human and the scope of being an animal thus turn out to be rather wide—they cover an infinity of things. Furthermore, the infinities appear to be of the same cardinality. If this is right, then it appears that the two properties have the same degree of determinacy and essentiality, which is precisely what the revised account wants to avoid.

Fortunately, this worry is not devastating to the account. Some philosophers have argued that infinities of the same cardinality can be compared. For example, Vallentyne and Kagan (1997) present some compelling examples that involve comparing infinite values. We might imagine two worlds that have a first time slice but no last time slice (i.e., the worlds extend infinitely into the future). Now imagine that at all times in the first world, the world has one unit of value. At all times in the second world, the world has two units of value. Adding the values at all the times for a world results in the overall value of the world. For both of these worlds, that value is infinite. Furthermore, those infinite values have the same cardinality. Nevertheless, it seems that the second world contains more value than the first. Kagan and Vallentyne propose establishing an isomorphic mapping between the times of the two worlds. If one world is such that each time slice has more value than its corresponding time slice in the other world, the world with the time slices that have more value than their corresponding slices in the other world also has more value than the other world.

The general strategy implemented by Vallentyne and Kagan (1997) is applicable to degrees of determinacy. Suppose that the two properties in question are being human and being an animal. Instead of counting up all the humans and animals in all possible worlds, look instead at the number in each world. For every world, the number of things that are humans is equal or less than the number of things
that are animals. Since humans are a subcategory of animals, it is impossible for the number of humans to exceed the number of animals. Even if there are infinite numbers of both in all possible worlds, we can use the numbers at each world to create an isomorphic pairing between the numbers of humans and numbers of animals at each world. Thus, even if there are infinite animals and infinite humans, it can still be shown that there are more things that are animals than things that are human. Consequently, the property of being an animal is less determinate than the property of being human.

It is also worth nothing that even if this solution isn’t applicable to all cases—perhaps it won’t work in the case of being a moral patient and being chordate—determinacy can still get give us a thing’s most determinate and least determinate properties. An individual’s most determinate properties are the ones that only it has. In the case of kinds, the most determinate properties are the ones that only that kind has. The least determinate properties of an individual or kind will be those that all things have. The cases in which it may be difficult to assess which of two properties is more determinate (either because we’re epistemically impoverished or because there’s no fact of the matter) are cases in between these two extremes. And since what matters for extensional similarity with the definitional conception is the thing’s most essential properties, as long as we can figure out what a thing’s most determinate properties are, the modification will have done the crucial work with which it was tasked.

With just this preliminary sketch, we can see how the revised modal account avoids the problems associated with non-discriminating necessary properties. It also avoids the problem of essential existence, assuming that existence turns out to be a trivially necessary property. Equipped with determinateness, the revised modal account begins to resemble the definitional account insofar as it can distinguish properties that seem to get at what a thing is from properties from properties that
all objects have. The most essential properties of a thing are the properties that definitional accounts would tend to say are essential. It can also deal with the disjunctive and negative properties that are necessary but seemingly inessential. For example, the property of not being Plato is shared by everything but Plato. It is thus not very determinate, and in turn not very essential to Socrates. The property of being a human or a mountain is slightly more essential than not being Plato, but not as essential as being human since the scope of what a thing with that property could be is still somewhat broad. By appealing to determinacy, the revised account gives the intuitively correct result that disjunctive properties are going to be less essential than the property involving just one of the disjuncts. In addition to the completely non-discriminatory necessary properties, the slightly more discriminatory properties that still seem inessential are also accounted for by determinacy. However, the challenge posed by coextensive and derivative properties still remains.

2.5.2 Stage 2: Sparseness

The second stage of revision, sparseness, addresses the challenge posed by coextensive and derivative properties. Several proponents of the modal conception have proposed revising it by restricting the necessary properties that are candidate essential properties. This revision based on sparseness has the following schema:

\[(\text{Schema}) \quad \text{a is essentially F iff (i) necessarily, Fa if a exists, and (ii) the property of being F is a sparse property.}\]

What counts as a sparse property differs wildly. Schaffer (2004) offers two conceptions of sparse properties—one that includes metaphysically fundamental properties and one that includes scientifically fundamental properties. Wildman (2013), picks up on the metaphysically fundamental conception and gives a sparse
modal account of essence on which sparse properties are the ones with metaphysical
priority. Denby (2014) offers a different account of sparseness on which intrinsic
properties are the sparse ones. There are other variations as well that analyze
sparseness in terms of non-triviality (Della Rocca 1996) or non-accidentalness
(Gorman 2005). Though these analyses of sparseness differ, they function similarly
once sparseness incorporated into a modal account of essence. There are fewer
essential properties than necessary ones. Thus, sparse accounts avoid the Finean
counterexamples based on necessary coextension and determinateness, such as
being the sole member of Socrates’s singleton set.

My analysis of sparseness departs from all of these. Like the other sparse
theories, I offer an account on which there are fewer candidate essential properties
than initially assumed. For instance, it also turns out that on my account, there is
only one essential property amongst being Socrates and being the sole member of
Socrates’s singleton set. However, unlike the other sparse accounts, I don’t draw a
distinction between sparse and non-sparse (or abundant) properties. Instead, there
is only one property under consideration in the first place. On my analysis,
necessarily coextensive properties are the same property. For convenience, let’s call
this claim the ‘necessary coextension thesis.’ According to the thesis, the property
of being Socrates and the property of being the sole member of Socrates’s singleton
set are necessarily coextensive, and thus turn out to be the same property.

I admit that this view of coextensive properties is somewhat counterintuitive. For
instance, triangularity and trilaterality are necessarily coextensive, and the former
appears to be a property about angles while the latter appears to be a property
about sides. But that they are in fact the same property has been defended by
several different philosophers. For instance, on the Lewisean account of properties,
properties are classes. To have a property is simply to be a member of a class. Where two properties are coextensive, then the corresponding classes contain the same members. Since members of classes wholly determine the identity conditions of their classes, coextensive classes (and thus properties) turn out to be identical to each other.

Oddie (2005) defends the necessary coextension thesis in a different way. On his view, there are distinct procedures or specifications of properties, but only one property where there is necessary coextension. For instance, just as subtracting two from five and taking the square root of nine are two different procedures for arriving at the number three, being the difference between two and five and being the square root of nine are two ways specifications of the property of being three. Similarly, trilaterality and triangularity are two specifications of the same condition—the condition had by all triangles. A procedure or specification might turn out to be prior to some other procedure or specification, but there is only one property in question. My goal is not to mediate between different views on why necessarily coextensive properties are the same property, but to show that if the thesis is true, then there is nothing objectionable about saying that the property is essential. If being the sole member of Socrates’s singleton set is the same property as being Socrates, and the latter is essential, then so is the former—after all, they are the same property!

The upshot of this analysis of sparseness is that it avoids several problems that its competitors encounter. First, as Skiles (2015) points out, on accounts where sparseness is analyzed in terms of metaphysical or scientific fundamentality, many seemingly essential properties end up being inessential. For instance, being human is usually taken to be an essential property of Socrates. But the property of being

---

10To be clear, Lewis (1983) distinguishes between properties and universals. The former are what I describe above. The latter are sparse, and comprise the minimal basis for characterizing the world completely. Some people refer to universals as “necessary properties”.

human is likely not metaphysically or scientifically fundamental, which thus rules it out as being essential. Problems also arise for accounts like Denby’s that are based on some other feature like intrinsicality. Some objects have essential relational properties; for instance, the Eiffel tower is essentially man-made or designed by so-and-so. Since my account does not rely on any sort of fundamentality, intrinsicality, or the like, it avoids these problems. Second, giving an analysis of sparseness that is not purely modal runs dangerously close to removes the main advantage of having a modal account in the first place. The main advantage of the modal conception is that it is reductive. But if one gives an account of sparseness that introduces new primitives—as, for instance, Wildman and Gorman do—then the account of essence is no longer reductive. My analysis of sparseness is purely modal. It thus is able to enjoy the advantage of being purely modal and does not introduce unnecessary primitives.

2.5.3 Summing Up

Now that the two stages of my revised account are in place, we can see that my revised account is a purely modal one that is extensionally similar to the definitional account. By using determinateness to order necessary properties in degrees of essentialness, the account can distinguish between more or less discriminatory necessary properties just as the definitional account does. Furthermore, via sparseness, it avoids the problem of necessarily coextensive and derivative properties altogether. My revised modal account is not perfectly extensionally similar to definitional accounts. As I’ll explain in the next section, this leads to two theoretical advantages. However, my account still is similar enough to do all of the same work that definitional accounts do. First, it yields similar results with respect to properties that are necessary but seemingly inessential. On my account, these seemingly inessential properties turn out to be
the least essential ones. Second, because we can talk about the most essential properties of a thing, we can enjoy the metaphysical explanation that definitional conceptions have to offer. For instance, the most essential properties are the ones that tell us about a thing’s nature. Third, my account individuates kinds and individuals just as well as the definitional account does. And finally, to the extent that the definitional conception is joint-carving, the revised modal conception is as well since it can approximate those joints.

2.6 Two Theoretical Advantages

Before concluding, I would like to point out two theoretical advantages that my revised account has and that the definitional account lacks. First, my account is reductive, and thus does not introduce any new primitives. Definitional accounts, on the other hand, either take essence as primitive or must use some other primitive such as fundamentality to distinguish between essential and inessential necessary properties. Thus, my account is more parsimonious. Insofar as parsimony is a virtue, my account has a virtue that definitional accounts do not.

Second, my account avoids arbitrariness when it comes to borderline cases. Plausibly, there are cases where it might be vague whether a certain property gets at what a thing is. For instance, being capable of moral reasoning seems to be one of the most significant properties that humans have. However, it is possible that this property does not rise to the level of essentiality for humans under the definitional account since it might follow from the property of being rational. On the definitional account, borderline definitional properties have to be either essential or non-essential. Classifying them as one rather than the other appears arbitrary. A similar problem arises in cases of necessarily coextensive properties. Without adopting the necessary coextension thesis, there can be two properties that seem like equally good candidates for essentiality. For instance, both
triangularity and trilaterality get at what a triangle is, and neither is obviously more fundamental than the other. But definitional accounts are not inclined to say that both triangularity and trilaterality are essential to triangles. Yet designating one rather than the other seems arbitrary in light of the fact that there aren’t good grounds for holding that one is more essential than the other.

On the revised modal account, we don’t need to make judgments in any of these cases. All necessary properties are essential, and then the property of being capable of moral reasoning is simply more essential than properties like being an animal and roughly as essential as being rational. Admitting of degrees of essentialness is more advantageous than attempting to draw a sharp cut off between necessary essential properties and necessary non-essential properties. It avoids the implausible result that there is a sharp cut-off between properties that get at what a thing is and properties that do not. While the definitional account is committed to this implausible sharp cut-off thesis, the revised modal account is not. The revised modal account also is not committed to distinguishing between necessarily coextensive properties, since, via sparseness, they are the same property. My account thus handles close cases better than the definitional account does.

2.7 Conclusion

These last two points show that the revised modal account goes beyond the definitional account. The revised account is extensionally similar to the definitional account, and the most essential properties of a thing can be called upon to do similar work to that of the definitional account’s essential properties. Furthermore, the revisions I have proposed are better at approximating the extension of the definitional account than those proposed by competing revised modal accounts. Finally, because my account is still purely modal, it preserves the main advantage of the traditional modal account—a fully reductive analysis of essence. Though
Fine’s counterexamples expose a flaw in the simple modal conception of essence, there is a revised modal account that is fine-grained enough to perform the work that we want essence to do. In developing my revised account and showing the work it can do, I hope that I have successfully revived the modal account of essence.

2.8 Cited Works


ESSAY 3

THE PROBLEM OF SELF-TRANSFORMATION

Abstract. From the standpoint of self-interest, is it rational for me to prefer to become someone who is radically different from my current self rather than someone who feels connected to who I currently am? Should my answer to this question depend on value considerations associated with the potential ways my life might go? These questions—which constitute what I call the problem of self-transformation—are the focus of my paper. In attempting to answer these questions, I hope to show (i) that value considerations underdetermine what preferences agents ought to have in decisions involving self-transformation, (ii) that considerations grounded in the will of an agent are needed to determine rational preferences in these decisions, and (iii) that there is a distinction between who agents are practically and who they are metaphysically with the former being what matters for self-interest.

3.1 Introduction

Suppose that I’m approached by an omnipotent being who makes me the following offer:

I can make you the happiest person in the world. In fact, I can guarantee that your happiness will be unsurpassable. The catch is that you will become radically different from who you are now. Your life goals, core preferences, and the way you see the world will be different. But rest assured, once you change, you won’t mind parting with all the things with which you most strongly, currently identify. Becoming the happiest person in the world will radically transform you.

From the standpoint of self-interest, is it rational for me to take the omnipotent being up on this offer?
Before beginning to answer this question, a few preliminaries are in order. First, I take this question to be about practical rationality, or what is good for the agent in question. I’m concerned with what agents should do from the standpoint of self-interest, not necessarily with what they ought to do from the standpoint of morality. Morality matters to the extent that the agent cares about being a moral agent and assigns value to moral outcomes. Relatedly, there are views on which agents can have reasons (such as, but not limited to, moral reasons) that are not self-interested reasons. On these views, it’s possible that what one ought to do all things considered comes apart from what one ought to do from the standpoint of self-interest. Again, I’m concerned purely with what agent’s ought to do from the standpoint of self-interest, not necessarily what they ought to do all things considered.

In addition, it is important to distinguish between the standpoint of self-interest and an objective, agent-neutral standpoint. For instance, consider the preference a self-interested agent might have when given the choice between an hour of pleasure and an hour of pain. Clearly, the self-interested agent would prefer the hour of pleasure. From the standpoint of self-interest, she would prefer the hour of pleasure even if that outcome resulted in a global state of affairs that was worse overall than the global state involving the hour of pain (e.g., even if everyone else in the world experiences an hour of pain in the outcome where she has an hour of pleasure and if everyone else in the world experiences an hour of pleasure in the outcome where she has an hour of pain). Self-interested agents are thus interested in value considerations that center upon them. Let us refer to these values as *agent-relative values*. Agent-relative values can be contrasted with *global values*, or the total value of a state of affairs that is accessible from an impersonal third-person perspective. A helpful heuristic for thinking about agent-relative and global values is to think of the values with which the egoist and the utilitarian are concerned. The egoist is concerned with agent-relative values while the utilitarian is concerned with global values.
Second, to keep things simple, I’ll assume that agents are correct in their judgments about what is good for them. This assumption is innocent, since practical rationality takes into account agents’ beliefs rather than what is true of the world. For instance, instrumental rationality is about what agents ought to do given their desires and beliefs as opposed to their desires and what is actually the case. Practical rationality is about what is internal to the agent.

Finally, and again for simplicity, I’ll assume that there is no uncertainty with respect to agents’ beliefs. For instance, in the example above, let us assume that I believe that the omnipotent being can and will carry through with the offer should I assent to it. As will become clear soon, I’m interested in the types of considerations that ground the rational preferences of self-interested agents, not which of the competing standard decision procedures ought to be implemented. In some sense, the issue I’m interested in is prior to questions about whether we should try to maximize expected value or go for the outcome with the best worst-case scenario. Part of what I’ll do is question whether an assumption about value considerations shared by standard decision procedures is one that ought to be made in the first place.

With these preliminaries in place, here’s a brief overview of how I plan on answering the question of whether, from the standpoint of self-interest, it is rational for me—or any other agent—to prefer and choose radical transformation. I start by suggesting that these cases involving radical transformation form a special class of decision cases that I call self-transformation cases (§ 3.2). Self-transformation cases appear in various forms throughout the literature on practical reasoning, so I situate my project against that backdrop (§ 3.3). The orthodox position in the literature operates on the assumption that agent-relative value determines what an agent’s self-interested preference ought to be. In the heart of my paper, I challenge this assumption and argue that in cases of self-transformation, value considerations
do not fully ground self-interested preferences (§3.4). This result leads to a problem: what, if anything, grounds an agent’s self-interested preference in cases of self-transformation? The remainder of the paper explores a solution based on non-evaluative grounds. I suggest that first-personal grounds—considerations that are grounded in the will of an agent rather than value—are needed in these cases (§3.6). This solution has some surprising implications for who we are metaphysically and practically speaking (§3.7).

3.2 Self-Transformation

The subject of this paper is self-transformation. To get clear on what it is is, let’s start with some paradigm cases and highlight their salient features. The first case is borrowed from Paul (2015a) and involves an agent who is deliberating about whether to have her first child. As Paul tells it, the potential parent faces multiple dimensions of transformation. First, becoming a parent involves *epistemic transformation*—the agent will learn what it is like to become a parent and the subjective value of that experience. Second, becoming a parent involves *personal transformation*—the agent’s core preferences, life goals, and way she sees and experiences the world may radically change. This personal transformation also contains an element of epistemic discovery since the agent will learn the ways in which she has transformed. More importantly, the personal transformation involves the agent actually changing, and doing so in a way that is potentially alienating. Conceptually, we can separate the epistemic changes (i.e., the discovering what the experience is like and how the agent will change) from the non-epistemic changes to the agent (e.g., changes to how the agent experiences the world, core preferences, and life goals). These non-epistemic

---

1In some case, the epistemic change can constitute the personal change. For instance, the agent’s knowing what it will be like to experience the world in a different way might be the change to how the agent experiences the world. In cases where there is some overlap between epistemic and personal transformation, I’ll count the change as personal since I’m interested in personal transformation.
changes are the ones that are relevant to self-transformation. From the perspective of the potential parent, the future parent might be so alien that the potential parent cannot identify with the future one in virtue of these non-epistemic changes.

The second case is inspired by Parfit (1984). It, too, highlights the alienation that can occur as a result of radical change. Parfit asks us to consider the case of a young Russian who holds socialist ideals. He holds these ideals so strongly that he regards them as “essential” to him. Losing those ideals would be tantamount to non-existence, and he expresses this to his wife when he implores her to prevent his future self from reneging on his plan to give away the vast estates he will inherit. The young Russian realizes that he may lose his socialist ideals when he ages, and he views that future, non-socialist self to be alien because it lacks the features of his current self with which he most strongly identifies. The disconnect between his current and future selves arises because of the loss of features that he regards as essential. Undergoing a transformation in which he loses these features, from his perspective, is as bad as non-existence.

In addition to these two examples, there are other paradigm cases of self-transformation. Paul (2014) gives us yet another example when she asks us to consider whether it would be rational to accept the offer to be turned into a vampire. Given that it would not be immoral to be a vampire and that ample testimony in favor of transformation from those who have become vampires, you might have strong reasons to believe that the experience of becoming a vampire would be highly valuable and that you would not regret it once you changed. Nevertheless, if you can’t identity with your potential future vampire self and regard it as alien, then it’s still not obvious that you should transform. Something similar holds in more mundane cases many of us actually face. For many, religious conversion or apostasy would involve radical changes akin to the change Parfit’s young Russian faces. It’s not hard to imagine Saul regarding his future self as the
apostle Paul as completely alien from his pre-converted self. Some people considering retirement report that part of their apprehension with respect to retiring is that not doing what they have done for so long is totally foreign to them. We can easily imagine a similar disconnect between an earlier and later self occurring when agents make major changes such as going to war, committing to marriage, or adopting a new vocation.

All of these cases involve agents who undergo radical change. These changes are not ones that call personal identity into question; instead, they involve changes to features that the agent considers essential to the agent’s self. The agent’s life goals, deeply held ideals, and particular way of experiencing the world are at stake. Again, these features are not essential in the metaphysical sense—otherwise the agent would cease to exist and be replaced by a new person should they be lost. But these features are treated as practically essential since losing them results in a future person who is disconnected and alien from the present agent.

It’s important to note that the presence of the type of change in question is not always sufficient for generating a case of self-transformation. For instance, we might imagine a young American who, like the young Russian, has socialist ideals but eventually will inherit a fortune and abandon his ideals. But unlike the Russian, the American is more apathetic about his ideals. He’ll vote for Bernie Sanders in the primary and be mildly disappointed when Sanders loses, but has other life projects that occupy the majority of his energy. He doesn’t regard his socialist ideals as essential to who he is, and thus does not view his future self, sans ideals, as alien. Whether a change involves alienation thus depends on what that agent treats as essential to him or herself as opposed to the change itself. In turn, whether a decision is one that involves self-transformation depends upon how the agent regards her potential future selves.²

²It’s possible for agents to be mistaken about whether a particular decision involves self-
In light of these considerations, we can define self-transformation as the change undergone by an agent, S, between times t and t’ when and only when S at t’ lacks some feature, F, that S at t regards as practically essential. In the cases of deliberation about the future, t is the time at which the agent is deliberating and t’ is the later time at which the future alien self is located. It’s also possible for t’ to be earlier than t. For instance, one who has undergone transformation may regard her past self as alien. Perhaps this is how many of us regard our past, newborn selves. Finally, being regarded as lacking a practically essential feature is not necessarily symmetric. For instance, from one’s current perspective, her transformed future self may lack a feature regarded as essential by the current self, but from the perspective of the future self, the earlier pre-transformed self may not lack a feature the future self regards as essential. All that we need for an interesting case of self-transformation is alienation in at least one direction.

3.3 Responses to Self-Transformation

Self-transformation is a phenomenon that appears in the literature on practical reasoning, though perhaps not by that name. It presents a special challenge to rational decision making because the connection between self-interested preferences and agent-relative values is not as straightforward as it is in ordinary cases. Recall that when faced with the choice between an hour of pleasure or an hour of pain, the self-interested agent prefers the hour of pleasure because it has more agent-transformation. For instance, one might not believe that having a child involves self-transformation and then realize after having the experience that it does. To a certain extent, one could research people who have undergone a particular kind of change and try to discover the percentage of people who report that the change is transformative. But for the sake of simplicity and given the context of decision theory, we can stipulate that agents are not mistaken since the agent’s beliefs are what are relevant for practical rationality. Even when agents are mistaken, there is still an interesting question about what the agent ought to do given her internal states and beliefs. (For instance, we evaluate whether agents are instrumentally rational by considering their desires and beliefs about how to satisfy those desires—not their desires and what actually would satisfy them.)
relative value than the hour of pain. That the self-interested agent ought to prefer
the state with more value for her is almost analytic. However, picking the outcome
with the most agent-relative value does not appear to be analytic in cases of self-
transformation. For instance, recent psychological research suggests that parents are
less happy than their childless counterparts. Suppose an agent deliberating about
whether to have a child knows that she experiences less agent-relative value in the
parenthood outcome than in the childless outcome. If having a child has always
been a core part of that agent’s life plan, is it really in that agent’s self-interest to
abandon the plan in light of the value considerations? And even if it ultimately is in
the agent’s self-interest to prefer the outcome with more value, why do cases involving
self-transformation feel different from ordinary cases?

Philosophers have offered various diagnoses of these cases that purport to
explain why they are different from ordinary cases. For instance, Paul (2014, 2015)
argues that in these cases, agents are not able to project what the experience will
be like. Transformative experiences, which include the cases of self-transformation
we’re interested in such as the potential parent one, have two elements—epistemic
and personal transformation. Agents who haven’t had the relevant experience are
not epistemically placed to know what the experience will be like or what they will
be like post experience. If agents cannot know what an experience will be like or
how it will change them, the value of the experience will be inscrutable. Without
the relevant values, we can’t form preferences about the outcomes involving
transformative experience, and standard decision procedures can’t be used. Thus,
decisions involving self-transformation differ from those involving non-transformative experiences because in the latter there simply are no values
upon which an agent can form a preference.

Paul’s picture is complicated because on her view, agents must be the ones to
assign the values. These assigned values—what she calls subjective values—are
similar to agent-relative values insofar as they represent the value the outcome will have for the agent. But even on this picture, we can conceptually separate the epistemic difficulty in assigning values from the transformation that occurs. Indeed, doing so reinforces Paul’s overall point that transformative experience defies standard decision procedures. Even if it’s possible for an agent to know the relevant values of the potential outcomes (and there are some persuasive arguments for thinking that we can), then the question of whether agents are rationally bound to prefer the outcome with the highest value still arises. While it’s true that value considerations underdetermine what agents rationally ought to prefer in transformation cases when values are inscrutable, when we stipulate what the values of the outcomes are, it still turns out that rational preferences involving self-transformation are not as obviously determined by value the way that ordinary preferences are.

Parfit diagnoses the problem in a different way. For Parfit, there’s no inscrutability with respect to value. Instead, the cases involving self-transformation differ from ordinary cases because they involve preferences about temporally distant selves. What ends up happening is that we discount the interest of our future selves, and thus discount the agent-relative value associated with future experiences. Given future discounting, the young Russian is rational in disregarding

\footnote{On Paul’s view, subjective values need to be assigned by the agent. I don’t adopt this assignment requirement for agent-relative values, but if one were to use subjective rather than agent-relative values, it would be trivially true that subjective values underdetermine what we ought to prefer in self-transformation cases since self-transformation cases are a subset of transformative experience cases and there are no subjective values to be assigned in transformative experience cases.}

\footnote{Much of the discussion on transformative experience has centered upon an epistemic battleground where fights tend to revolve around whether agents can know something about what paradigm cases of transformative experience are like without experiencing them and whether lacking knowledge of what the experience is like precludes the possibility of assigning values to outcomes. For example, Harman [2015] takes up the former issue and suggests that caring for a younger sibling in a quasi-paternal way and witnessing the relationship between a close friend and her first baby provide at least some evidence of what it is like to be a parent. Dougherty et al. [2015] address the latter point and argue that direct experience is sufficient but not necessary for assigning subjective value to an experience. Testimony, for instance, might provide an alternative route.}
the value his future non-socialist self will experience. However, temporal distance cannot be the full explanation for why cases of self-transformation do not obviously require agents to prefer future outcomes with more value. If we alter the case so that the young Russian will become wealthy and lose his ideals the next morning, his concerns about losing a practically essential feature will still apply. We could also alter the case so that the young Russian has a choice between two futures: one in which he inherits the money and loses his ideals and one in which he inherits the money, keeps his ideals, and gives the money away. If we stipulate that he experiences higher agent-relative value as a member of the economic elite than he does as a struggling socialist, it’s still not obvious that he should opt for transformation in virtue of the value considerations. Thus, temporal distance and future discounting does not explain why what one ought to rationally prefer is not as obvious in self-transformation cases as it is in ordinary cases where value determines rational preference.

Neither the Paul nor the Parfit diagnosis fully explains why preferences regarding self-transformation differ from ordinary preferences. In ordinary cases, our rational preferences are obviously determined by agent-relative values. But in cases of self-transformation, it is not obvious that agent-relative values determine the preferences we ought to have. In the next section, I’ll offer my own diagnosis of this situation. It’s not obvious that agent-relative values determine the preferences we ought to have in self-transformation cases because agent-relative values don’t determine the preferences we ought to have. Self-interest permits agents to prefer a potential outcome with less agent-relative value than an alternative potential outcome. Because agents’ rational preferences are not bound by value considerations, value alone is not sufficient for determining what agents ought to prefer and, ultimately, choose. We’re then left with the question of what, if anything, grounds rational preferences regarding self-transformation. Thus, the
problem of self-transformation arises.

3.4 The Problem of Self-Transformation

My goal in this section is to argue for the claim that when agents are deliberating
about whether to undergo self-transformation, value considerations do not fully
ground what agents ought to rationally prefer from the standpoint of self-interest.
The truth of this claim leads to the problem of self-transformation. But before
getting to the problem, let’s turn to my main argument:

(P1) In self-transformation cases, the rational preferences of self-interested agents
with respect to their pasts are not fully grounded in value considerations.

(P2) The grounds for rational preference are insensitive to whether the preference is
about the agent’s past, present, or future.

(C) In self-transformation cases, the rational preferences of self-interested agents
are not fully grounded in value considerations.

Both premises need explanation and defense. I provide both below. After (hopefully)
showing that the argument is sound, I’ll explain how its conclusion forces us into the
problem of self-transformation.

3.4.1 Defense of (P1)

In order to motivate (P1), I’ll start with a pair of cases. The first is a variation
on the non-identity problem:

Non-identity. Wilma wants to have a child. The doctor tells her that
she can have one now or take a pill and wait a month. If she waits, the
child will have LeBron’s physical gifts, Parfit’s philosophical ingenuity,
and Mother Theresa’s kindness. If she conceives now, the child will be
average with respect to physical and philosophical abilities, as well as
altruistic inclinations. Wilma conceives now, and Wilma’s child considers
whether it would be better for her had her mother waited.
Even if we stipulate that the world, from an impersonal neutral perspective, would be better had Wilma waited to conceive and had the second child, it would be irrational for Wilma’s child to wish that her mother had waited. After all, had her mother waited, she would not exist at all! It would not be in her self-interest for her mother to have waited. What this case shows us is that global value does not ground rational self-interested preferences. Though global value might be sufficient for grounding a third-person preference it isn’t sufficient for grounding a self-interested preference for Wilma’s child. Global value doesn’t fully ground self-interested preferences. Agent-relative values are needed.

Our second case, Past trauma, involves thinking about what would have occurred had the past been different:

*Past trauma.* Jon suffered childhood trauma. He eventually overcame those experiences such that those experiences shape who Jon has become. We can imagine an alternative past in which Jon’s early life went better and did not contain that trauma. Further, we can imagine that the better past does not detract from the value of the later stages of his life. Thus, non-traumatized Jon would have a life that contains more value in the past and overall, though he would also be a very different person from actual Jon. Jon now considers whether he prefers that his early life had gone better even though it would result in his being radically different from who he is.

Jon’s life would be better if his past had gone differently. There is greater agent-relative value in the alternative past scenario. Nevertheless, it doesn’t seem that Jon is rationally required to prefer that the past had gone differently. It is not hard to imagine Jon reasoning as follows: *it is true that had my early life gone better, my life overall would be better; however, I would be radically different than I am now, and I don’t want a life so alien from my own.* There is something persuasive about this kind of reasoning, and if it’s rational, then agent-relative value underdetermines what our past preferences ought to be since it can be rational to prefer the past with less agent-relative value.
Jon’s reasoning is persuasive because Past trauma is similar to Non-identity. In both cases, an agent compares two scenarios, one that involves who they are, and one that involves an alternative scenario with a radically different agent. In Non-identity, the personal identity relation does not hold. In Past trauma, there is personal identity, but there’s been significant divergence. This divergence involves changes to features that Jon considers essential to who he is. The way in which he overcame his past trauma has been incorporated into how he identifies as an agent, and constitutes who he is. When he considers the alternative non-traumatized life, he can’t identify with that alternative agent because that agent’s life is so different from his own. As far as Jon’s practical reasoning is concerned, that agent may as well be someone who is not personally identical to him. Thus, in that alternative scenario, from Jon’s perspective, Jon may as well not exist.

The lesson I’m drawing from this case is not idiosyncratic. In recent work, ? distinguishes between affective and cognitive empathy. When one affectively empathizes, one imagines what it would be like for them to be in a particular situation. Jon, when he affectively empathizes with non-traumatized twin-Jon, imagines what it would be like if he were to occupy twin-Jon’s position. This affective empathy is different from cognitive empathy, which requires Jon to understand twin-Jon’s perspective well enough to imagine what it would be like from twin-Jon’s subjective perspective. In general, it’s easy to cognitively empathize with our near past and future selves because our current perspective is relevantly similar to the perspective occupied by our temporally near selves. But cognitive empathy becomes more difficult the further into the future our future self is. And if the future self is located after a transformative event, cognitively empathizing with that self may be akin to empathizing with a completely different
The same applies for alternative possibilities. Where the alternative is too distant, as is the case with Jon and twin-Jon, cognitive empathy isn’t possible. Even if the alternative contains a metaphysical counterpart, the gap between the actual person and that counterpart is as wide as the gap between the actual person and a metaphysically distinct person.

Adams (1987) also holds a similar position. In the context of considering the problem of evil, he asks whether we owe our existence to past evils. For Adams, this question extends past the non-identity cases to cases where personal identity is not in question:

Even if I could, metaphysically or logically, have existed without more past evils and their consequences in my experience, I doubt that that existence could have been mine in such a way as to matter much from the point of my self-interest because it would not bear... “the self-interest relation” to my actual life. (68)

Adams’s paradigm case is the life of Helen Keller. When we imagine what her life could have been like had she not become blind and deaf, we can easily see that it would have been radically different. Even if the alternative life had contained more value, it seems perfectly reasonable for Keller to prefer her own life and the distinctive way that it turned out. Parfit (1984), too, acknowledges that what matters to us for survival is not quite personal identity. One of the lessons of the young Russian is that what matters is connectedness. Though the young Russian is the same person as the...

Velleman (1996) makes a similar point when he argues that when we consider what it takes to have a survival that matters to us, it is essential that we occupy the right point of view. Jon and twin-Jon do not occupy the same point of view, and from Jon’s perspective, twin-Jon does not bear the right relation to him for self-interest to hold.

By ‘counterpart,’ I mean someone who has the same metaphysical profile as the agent in question. For our current purposes, this could be analyzed in terms of sharing the same essential properties or Lewisian counterpart theory. Incidentally, this point about distant counterparts provides a way for the proponent of counterpart theory to respond to Kripke’s argument from concern. According to this argument, it’s paradoxical that Humphrey cares that he could have been president but not that one of his counterparts is the president if the former reduces to the latter. But we shouldn’t expect that Humphrey would be concerned for his counterparts (even if they were identical to him and not merely similar) because he can’t cognitively empathize with them.
future wealthy non-socialist, the young Russian is disconnected from that future self because it lack his features about which he cares most. His other thought experiments purportedly demonstrate this exact point.

Just as global value does not determine the rational preferences of Wilma’s child, agent-relative values do not determine the rational preferences of Jon, Keller, or Parfit’s young Russian when they compare their lives to ones that are radically different. What these considerations suggest is that in cases of self-transformation, appealing to agent-relative values is defective in the same way that appealing to objective values is defective in non-identity cases. Self-transformation cases are relevantly similar to non-identity cases. What matters to the agent with respect to self-interest is lost in alternative scenarios. The agent doesn’t exist at all in the non-identity case, and the features the agent identifies most strongly with and considers practically essential to themself are lost in the self-transformation cases. Thus, agent-value considerations underdetermine what agents ought to prefer in self-transformation cases, and (P1) is true.

One might object to my defense of (P1) by questioning whether it is possible to form preferences about the past. One source of discomfort may arise from the fact that the intrinsic features of the past are fixed so that one cannot actually choose the preferred past states. But in general, we don’t need to be able to bring about a state of affairs in order to prefer it. For instance, I prefer world peace to states of war. I can’t bring about my preferred state—though if I could, I would—but there is nothing incoherent about me having that preference and not being able to bring it about. A related worry questions whether we can have preferences about the past.

---

7 Though we can’t change the intrinsic features of the past, we can change the relational ones. But if we’re concerned with choosing between the relational features of the past, this sort of discomfort won’t arise.

8 What would be irrational is if an agent’s choice did not line up with the agent’s preference given that the choice is something the agent could bring about. In cases where the agent cannot bring about the preferred state, a disposition to choose would suffice for the agent having consistent
since we know what the past is like and that it is impossible to change its intrinsic features. However, it seems possible to form preferences about the past even when we know its character. For example, when I check whether my lottery ticket has won, I prefer that it became the winning ticket both before and after I’ve checked what the winning numbers are.

A final source of discomfort might be that we cannot project what an alternative past would be like and thus cannot form preferences about our actual past and alternatives. There are some compelling arguments for why alternative pasts aren’t projectable, but I’m going to set those aside for the purposes of this paper. Even if those arguments are right, it’s possible that we might know the agent-relative values associated with alternative past states of affairs. For instance, suppose that the value of one past state has a negative valence while the value of an alternative past state has a positive valence. Further suppose that self-transformation is not an issue in either of these cases; perhaps these past states differ only with respect to what the agent has eaten for dinner. In this case, even though the agent cannot project what either state is like, the agent can rationally prefer the past with the value that has positive valence. Thus, the inability to project what an alternative past state of affairs would be like does not, in principle, make it impossible to form rational preferences about alternative past states of affairs.

3.4.2 Defense of (P2)

In cases of past preferences, agent-relative values are not enough to ground self-interested rational preferences in cases of self-transformation. I now want to turn to whether this point generalizes to preferences with respect to all times. If so, then preferences and dispositions to choose. For the purposes of this paper, I’m not going to discuss the irrationality that arises from a mismatch between preference and choice (or disposition to choose).

\[^9\text{See Phillips.}\]
agent-relative values will not be enough to ground self-interested rational preferences with respect to potential past, present, and future self-transformation. *A fortiori,* agent-relative values won’t fully ground self-interested preferences in the future cases we encounter in everyday life and with which we’re concerned. In this subsection, I’ll explain why it seems likely that (P2) is true and then defend it from some objections.

*Prima facie,* it seems that whatever grounds rationality is the same for preferences about the past, present, and future. (P2)—that the grounds for rational preference are insensitive to whether the preference is about the agent’s past, present, or future—is a reasonable default position. There are a variety of arguments defending treating past, present, and future preferences similarly. For instance, the Epicureans thought that temporal differences between preferences were arbitrary differences that rationality ignores. Greene and Sullivan (2015) defend this line by pointing out that when agents treat past preferences differently by discounting them, agents can be dutchbooked, which demonstrates that their preferences are inconsistent, and thus irrational.

Interestingly, these arguments show that rationality requires *weighing* past, present, and future preferences equally. Weighing these preferences equally is logically compatible with the grounds for preferences changing over time. One might, for example, hold that past preferences are grounded in objective values and present preferences are grounded in agent-relative ones. But this position is also

---

10 On the Epicurean view, self-interest is a matter of pleasure—the self-interested agent ought to maximize the amount of pleasure in his or her life. We can think of the amount pleasure an agent experiences during any given event as the subjective value that event has for the agent. What the Epicureans realized was that tradeoffs in pleasure often have to be made between near- and distant-future events, and that one’s life ends up having more total value when near- and distant-future events are weighted equally. See also Finocchiaro and Sullivan (2016).

11 Sullivan and Greene are responding to Parfit. Parfit presents thought experiments that purportedly show that it is rational for agents to prefer that unpleasant events be in their past rather than their future. Sullivan and Greene point out that assigning any kind of discount function to the past, one becomes vulnerable to dutchbooking. Because dutchbooks occur when there are inconsistent value and probability assignments, avoiding dutchbooks is often taken to be a necessary condition for rationality.
implausible. An agent who has a preference at a present time $t_1$ would have to change that preference as time passed and the present time became $t_2$ since the preference at $t_1$ would now have different grounds. But it is hard to see why this would be. Presumably, those objecting to (P2) will have to give some explanation for why past, present, and future preferences are not all uniformly grounded. Perhaps there is some asymmetry between the past, present, and future that makes a difference for preference formation. I’ll briefly consider two here.

First, one might hold that the past is importantly different from the future because while the past is fixed (at least with respect to its intrinsic features), the future is open. This metaphysics underlying this point is controversial, but I’ll grant it for the sake of argument. While we can potentially exercise control over the future and which states of affairs are brought about, we cannot exercise control over the past. This difference prevents us from making choices about the past. Perhaps it also prevents us from having preferences about the past. If this is correct, we won’t be rationally required to have particular preferences about the past. For instance, Jon isn’t required to prefer that his earlier life went better despite the higher subjective value associated with that potential past. But when making decisions about the future, Jon is required to prefer the outcome with the highest level of subjective value because he has control over bringing about that future state. Thus, (P2) is false.

This worry is tied to the discomfort that some have with past preferences. But as we saw in the previous subsection, it is possible to have preferences about states even in cases where we cannot choose those states. Similarly here, a lack of control amounts to the relevant choice being unavailable. We can’t make choices with respect to the intrinsic features of the past and we presently have no control over those. However, the same is true of some future states (e.g., world peace) over which we have no control and cannot choose. Even synchronic cases about the present can feature a
lack of control coupled with the ability to form a preference. For instance, given that I’ve purchased a lottery ticket, I have no control over whether I win, but I can prefer that I do. The lack of control does not affect the ability to form a preference. Though the asymmetry with respect to control might affect which options are live—in one case I won’t be able to do anything to make it the case that I experience more subjective value for breakfast and in the other I will—it does not affect my ability to form rational preferences.

Second, one might object to (P2) because of a causal asymmetry between past and future states of affairs. Past states determine future states, but not vice versa. Like the first objection, this one rests upon metaphysical claims that I will grant for the sake of argument. If we take (P1) seriously, one explanation for why Adams and I are correct in our assessment relies on the fact that the past determines the future. Jon, who is happy with who he has become, cannot prefer a significantly different past because that would entail radically changing who he has become. However, what happens in the future has no causal effect on who he is now. Thus, the objection goes, though one may not prefer a past that is significantly different to the one that led to the present, one can in principle prefer any of the potential futures since none of those can change who one is in the present.

In responding to this objection, it’s important to return to why Jon prefers his actual past involving the trauma to the one in which he does not experience the trauma. Jon doesn’t prefer his actual past merely because it is the only past that could result in who he is; he prefers it because he regards it as an integral part of who he is. He self-identifies with that particular past and is more than causally related to it. His past is constitutive of who he is now, and it is this connection rather than the causal one that allows him to form the preference for this past. In Parfitian terms, we might say that Jon’s past and his current relation with it is an essential feature who Jon is. We could also adopt the distinction that Schechtman
(1996) draws between ‘reidentification’ and ‘characterization.’ The former has to do with features that matter for metaphysical identity, while the latter has to do with features that are “truly those of the person” (73). The features that characterize are those that constitute who Jon is. To be sure, the causal and constitutive aren’t easily pulled apart, but carefully thinking about pairs of cases can help us see that this distinction does in fact exist and is significant.

Imagine twin-Jon, who is like Jon in every respect except for the way in which he connects to his past. Jon’s past has informed the person he has become, and is thus constitutive of the person into whom he has developed. While twin-Jon has followed the same path as Jon, he resents his traumatic past and wishes that it had not happened. Like Jon, twin-Jon likes the person he has become, would not necessarily prefer to be someone radically different, and realizes that his past has causally led to who he is. However, unlike Jon, twin-Jon resents his past. If there were a miracle world in which he developed into the same person without the traumatic past—a world in which the causal connection is severed—twin-Jon would prefer that world to the actual one. This shows that twin-Jon prefers his own past to competing alternatives in virtue of the causal connection, whereas Jon prefers his past in virtue of it being constitutive of who he is.

This distinction reveals why the causal asymmetry objection appears convincing. If we have in mind preferences that are based on merely in the causal connection, then (P2) appears false since the causal connection does not apply to future states. However, if the preference is based in the past’s being constitutive or essential to who the agent is, then this objection doesn’t demonstrate the falsity of (P2). Instead, the objection would need to rely upon an asymmetry between the past and future when it comes to being a constitutive or essential part of who a person becomes. But once we shift from causation to constitution, it is not clear that there is an asymmetry between the past and the future. Instead, if a potential future lack a feature that
is constitutive of the agent, then it seems that the agent is rational in avoiding that
future even if it is more valuable than one for which she ultimately opts. After all,
if a constitutive part is an essential feature of an agent, it’s essential for all times
the agent exists. So unlike causation, constitution does not appear to give rise to an
asymmetry. Instead, considerations with respect to constitution seem to give a very
good reason for thinking that future preferences work similarly to past ones. Thus,
the objection to (P2) fails.

Though this objection to (P2) is unsuccessful, it does show us something
important about the problem of self-interest. The asymmetry does not show us that
future self-interest is always grounded in agent-relative value while past preference
might not be in special cases like self-transformation. However, it does suggest that
while causal asymmetry might fix which pasts are alien to our current selves, it does
not fix which futures are alien to our current selves. This difference is important
not because it makes subjective value the sole grounds of self-interested preferences,
but because it suggests a possible solution to the problem of self-transformation to
be discussed in §3.4.3.

3.4.3 (C) and the Problem of Self-Transformation

The truth of (P1) and (P2) entails the truth of (C), that in self-transformation
cases, the rational preferences of self-interested agents are not fully grounded in value
considerations. Non-identity shows that global values don’t matter for self-interest.
Though agent-relative values determine self-interested preferences in ordinary cases,
Past trauma shows that they don’t for self-transformation cases. Given (C), a fortiori,
rational preferences aren’t fully grounded in value considerations in the future cases
of self-transformation that we encounter. From this point, it is easy to see how
the problem of self-transformation arises. Knowing all the relevant value facts does
not produce a unique rational preference, even if there is a unique, most valuable
outcome. Value underdetermines what one’s rational self-interested preference ought to be.

In some sense, this conclusion is not completely novel. We can imagine a modified version of Parfit’s case where the young Russian is choosing between two futures, one as a struggling socialist and one as a member of the bourgeoisies. Here, we can imagine that the young Russian knows that once he is a member of the bourgeoisies he will enjoy substantially more agent-relative value than he would if he were to remain a struggling socialist. Still, it will be the case that regardless of how much more value the future as a member of the bourgeoisies holds, the young Russian will still be alienated from it because it lacks the feature he considers most essential to his self. He will still regard that future as being as bad for him as a future in which he ceases to exist. Facts about value will not affect this judgment.

Paul also holds that agents cannot decide to undergo transformative changes on the basis of agent-relative values. Furthermore, even if agents are able to know the relevant values and adopt an observational perspective where they take themselves to be in the shoes of the transformed self, that will not provide them with the relevant information to make a rational decision. Agents also require “de se” knowledge, or knowledge that the transformed future person is in fact them. Prospection, or cognitively empathizing with one’s future self, involves this de se belief. We can do this in ordinary cases that involve our near future selves. However, the nature of transformation makes this type of prospection impossible. Without the relevant de se knowledge, there is no guarantee that the agent’s current, subjective self endures through the transformation, and value won’t be the thing that makes it rational to risk the survival of the subjective self.

Since value considerations underdetermine rational preference, there are then two options: either rational preferences in self-transformation cases are grounded in non-

---

evaluative considerations, or there are no grounds for rational preferences. In the latter option, preferences are then arational. It would be rather disturbing if our most pivotal life decisions could not be made rationally. We should only settle for this option after searching for other potential grounds. As I’ll argue in the next section, there won’t be a need to settle, because there are other grounds. The key in uncovering them is to turn away from value considerations, and look to the non-evaluative.

3.5 Solving the Problem: First-Personal Grounds

In order to solve the problem of self-transformation, we need something other than value to ground preferences. At first glance, this proposal may appear deviant or misguided. Our standard decision making procedures are built around value and don’t take non-evaluative facts into account. Even moral considerations are accounted for by objective value and can factor into subjective value in cases of agents who care about morality. The non-evaluative is not typically taken as a relevant consideration when figuring out whether a preference is rational. Thus, what I’ll first try to do is motivate the case for non-evaluative grounds for rationality. I’ll give three cases in which the most fundamental grounds for rational preferences are non-evaluative. I’ll then explain what these cases have in common, and argue that they strongly suggest that in special cases like self-transformation, rational preferences can be grounded in ‘first-personal grounds,’ or grounds arising from an exercise of agentive power, rather than value facts. Finally, I’ll consider and address three objections to this proposal.

3.5.1 Three Cases

(i) Performative acts. We can give ourselves reasons to act. When newlyweds vow to love and cherish each other for better and worse for the rest of their lives, they give themselves a new reason to continue to love and cherish each other. Pre-
vow, the couple might stay together because doing so is in their self-interest, but post-vow, there is a new practical reason to stay together. Staying together upholds their promise, which has value. There is disvalue to breaking the promise (though in some cases that might be outweighed by other considerations). What’s important about this case is that the value of keeping and breaking promises arises only as a result of the agents making the promise in the first place. It may appear as though keeping the vows are rational because that maximizes value, but on closer inspection, the more fundamental ground for why it is rational to keep the vows is located in the act of promising. The rationality of keeping their vows is traced back to decisions on the part of the agents, which are non-evaluative facts.

It is, of course, possible that the rationality of making the vows in the first place could be grounded in value. Perhaps one of the newlyweds is marrying for money or political power. In these cases, the rationality of the act would be grounded in value considerations, and the most fundamental grounds would still be evaluative. And in fact, many promises—imagine the typical contractual promise between a buyer and seller—are made because there is more value to making the promise than not. However, it certainly seems possible that in at least some cases, evaluative considerations underdetermine whether an agent ought to make the promise. In the case of marriage, perhaps the agent would do just as well as a single person or has equally good spousal candidates. Where value does not determine whether it is rational for the agent to make the promise, we are forced to accept one of two options. Either there is nothing grounding the rationality of making the promise, in which case doing so is arational, or there is something grounding the rationality of making it, in which case that something must be non-evaluative. I won’t argue for this point, but it certainly seems plausible that there are cases in which evaluative considerations underdetermine whether it is rational to make a promise, and that it is nonetheless rational to make these promises. As long as these cases are possible,
then there can be cases where the rationality of an action is not grounded in value.

(ii) Revelation. Paul’s own solution to the transformative experience puzzle involves revelation. On her picture, an agent considering whether to undergo or avoid a transformative experience cannot use value to ground the rationality of her choice since the transformative nature of the outcome precludes assigning the relevant values for self-interest. Thus, the agent must make the choice on different grounds if the choice is to be made rationally. Paul suggests that agents choose to discover what the experience will be like, and that doing so makes the decision to undergo the transformative experience rational. Importantly, there is no hidden value within discovery that makes this choice rational. Agents can also choose to avoid discovery, and be rational in doing so. Antecedent to the agent’s choosing to discover what the new experience is like, there is no extra value that discovery has over refusing discovery. Again, it may be true that given that the agent chooses discovery, the outcome with discovery has more value and ought to be preferred. However, the agent’s choice is what grounds any value discovery may have, so the agent’s act is again the more fundamental ground for rationality. Thus, we again have a case where an agent’s action, rather than an evaluative fact, grounds what the agent rationally ought to prefer.

(iii) Voluntarism. Ruth Chang has argued that we sometimes face difficult decisions where potential outcomes are ‘on a par.’ For instance, an agent might be choosing between being a professional philosopher or a professional athlete. Neither option is better than the other—if so, one would pick the better option and the decision would be easy. But the options are not perfectly equal either. If so, increasing the salary of or adding some other small benefit to one of the careers would tip the scales in favor of that one. Since the better than, worse than, and equal to relations do not hold between the two outcomes, our given reasons, which include evaluative considerations, underdetermine which outcome an agent ought to
prefer. As in the case of promises, it would be disappointing if there were no way to rationally choose one of the options in such major life decisions.

Chang’s solution is to appeal to voluntarist reasons—reasons that arise from the willing of an agent. By putting herself behind a reason in favor of one of the outcomes, an agent can strengthen that reason and tip the scales in favor of that outcome. For instance, if our agent choosing between professional philosophy and athletics wills that knowledge matters, she might tip the scales in favor of professional philosophy. The act of willing creates more reason, thus revealing that it was in fact rational for the agent to prefer it. When this happens, something an agent does, rather than evaluative facts, grounds the rationality of that preference. Korsgaard, like Chang, also defends the ability of agents to exercise agential powers in this way. For Korsgaard, agents create new reasons ex nihilo. When added to the evaluative considerations, these new reasons break the underdetermination and make it rational for the agent to do what the new reasons weigh in favor of. For our purposes, it won’t matter which of these accounts of voluntarism is correct. What’s important is that if there are voluntarist reasons, they can provide non-evaluative grounds for rationality in cases where value underdetermines what it is rational to prefer.

3.5.2 First-Personal Grounds

These three cases share two features. First, they involve choices in which value considerations underdetermine which outcome the agent ought to prefer. Second, in each of these cases, the agent facing the choice nevertheless manages to form a rational preference by acting. The agentive action, rather than value, grounds the rationality of the preference the agent forms (and in turn the choice the agent ultimately ends up making). For ease of reference, let us refer to these agentive acts as ‘first-personal grounds.’ These grounds constitute a class of considerations that are relevant to rational decision-making procedures. Crucially for our purposes, they provide a way
for agents to deliberate rationally about decisions involving self-transformation and yield a sufficient explanation for why the preference the agent ultimately ends up forming is rational.

Let’s return to our paradigm cases of self-transformation: Paul’s potential parent is deliberating about whether to have children, and Parfit’s young Russian contemplates a future in which he loses ideals and thus what he regards as most essential to his self. Grant that both of these agents can project what the potential future outcomes will be like and know the agent-relative value to those outcomes. The problem of self-transformation arises because these agents face a decision to radically change who they are. Having the agent-relative values is not sufficient for making the decision rationally. What also matters for agents are their core preference, life goals, and the way they see the world.

The addition of first-personal grounds to the decision-making calculus makes it possible for these agents to rationally decide to undergo or avoid transformation. On Paul’s story, potential parents do exactly this when they decide to become parents because they want to discover what it will be like. On a minimalist reading of Paul’s solution to the transformative experience puzzle, the potential parents are discovering what it will be like in the epistemic sense. But a more robust application of this solution—and this is what I take Paul to intend—helps with the self-transformative element as well. An agent might decide to undergo transformation to see what it will be like to transform. Alternatively, and agent might decide to avoid transformation because transforming would destroy who the agent is in a way that is deeply important to her. Furthermore, neither of these alternatives are open to rational criticism if they are opted for in virtue of the agent’s first-personal grounds. Indeed, it is the agent’s creation of these grounds that makes the preference rational.

Similarly, in the modified case with Parfit’s young Russian, we can picture the youthful socialist imagining what it will be like to lose his ideals and join the class
he despises. The young socialist might decide that he would like to discover what transforming into a member of the economic elite would be like, and do nothing to bind himself. Or he might put himself behind hedonistic reasons, will that those matter, and thus add more weight to a reason that would result in allowing himself to become an aristocrat. In willing this, he would have demonstrated that preferring to change is rational for him. Alternatively, the young socialist might resist transformation despite the higher subjective value and create first-personal grounds that make it rational for him to prefer to stay the same. Any of these decisions are rational once the first-personal grounds associated with them are added to the picture.

Besides providing the grounds for rational preferences involving self-transformation, first-personal grounds are especially important for these types of choices. In virtue of the first-personal grounds they create, agents end up connected in the correct way to their future selves. Because the ground of rationality arises from their agency, the future selves are no longer alien to the agent even if they are incredibly different. For instance, should the young socialist create first-personal grounds in favor of becoming wealthy, that future, in which the young socialist becomes radically different, would be connected to the present agent. Willingly abandoning his ideals would make it the case that they were not in fact essential to who he is, and thus make it rational to prefer to change in such an extreme way. Similarly, the present potential parent who chooses to discover what it will be like to have a child is connected to the future parent. We can see the connection when we contrast this case to one in which parenthood is thrust upon the agent. Even if the resulting parental experience is the same with respect to value, there is the possibility of a disconnect between the pre-parent agent and the parent in the case where the agent does not choose parenthood. This possibility does not arise in the case where the agent chooses parenthood. The use of
first-personal grounds connects the pre- and post- transformation selves such that the subjective perspective of the agent endures through the transformation.

This point also explains why agents can feel more or less connected to their past selves. Recall Jon and twin-Jon, who experienced childhood trauma. For Jon, the traumatic past constitutes who he is, and he would not prefer that it had not occurred. But for twin-Jon, the traumatic past causes, but doesn’t constitute, who he is. Thus, twin-Jon would prefer that it had not occurred. The difference between Jon and twin-Jon is that Jon has responded to his past in a way that integrates it into who he has become. Twin-Jon has not. Further, neither has responded to the past in an irrational way. Instead, one person did something that the other did not, and that action is what explains why it’s rational for one person to prefer their past over alternatives and why it is rational for the other to not have that preference. Again, an agentive act, or the creation of a first-personal ground, explains why the preference with respect to transformation is rational and how radically transformed agents can be connected to their pre-transformed selves.

3.6 Objections and Replies

Now that I’ve sketched my solution to the problem of self-transformation, I’ll address some potential worries about it. There’s no doubt that agents can exercise their agency. Our three examples of first-personal grounds involve ordinary acts that we perform: we make promises and thus create obligations for ourselves; we choose to undergo experiences to discover what they are like; and we put ourselves behind reasons such that they matter to us and then are guided by them. Thus, objections to the solution aren’t going to be about the existence of these grounds, but about whether they do in fact matter for rational decision making in the way I have suggested.

First, one might worry about whether first-personal grounds are created
independently of value considerations. For instance, take the case of the agent using voluntarist reasons to pick between careers in philosophy and athletics. Perhaps the agent willing that knowledge considerations matter reveals that the agent has a higher-order preference for knowledge. Ordinary preferences fall in the domain of value considerations—an agent’s preference for knowledge leads to a higher subjective value assignment to scenarios involving knowledge. Similarly, if there is a higher-order preference for knowledge, that preference affects the value assigned. If this is the correct diagnosis of the situation, then willing independent of value considerations is illusory. What’s really going on is that in deliberating, the agent has found additional value considerations on which to ground the decision.

While the lure of hidden value considerations is tempting, this epistemicist route is not viable in a range of cases, including those involving self-transformation. First, this route should only be tempting to those who think that an agent’s actions are determined by his reasons. Hume, for instance, held that in cases where it appears that agents act without reason, there are unknown reasons at work that ground the agent’s action. But if it’s possible to act without or against reason, then the fact that an agent performs an action after deliberating does not entail that the agent performed the action as a result of finding a reason in favor of it. Second, regardless of how one comes down on the first point, it’s surely possible that value considerations, hidden or apparent, can be perfectly balanced such that a deliberating agent who opts for one action or another does so without having found a value consideration that tips the balance. Cases in which options are on a par also exhibit underdetermination of value considerations. And, of course, cases, such as self-transformation, in which value considerations are irrelevant aren’t going to contain hidden value considerations that drive the agent’s acting. In these scenarios, there are no hidden value considerations that motivate the decision made. Thus, it is possible to act without being constrained by value considerations. It is in turn
possible to create first-personal grounds without them being grounded in value.

Second, if the creation of first-personal grounds is not constrained by value considerations, then the range of grounds that an agent can create appears unrestricted. We then might wonder why an agent creates one ground rather than another. This is particularly worrisome since if there is no explanation for why one ground is created rather than another, it seems that agents are acting arbitrarily. And in the context of decision-making, this result is disastrous since arbitrariness stands in tension with rationality. If there is nothing that explains why an agent creates one ground rather than another, then it seems that the agent should be indifferent between which ground to create. In turn, it’s hard to see how the agent is acting rationally when creating that ground.

In addressing this worry, it might be helpful to consider why it is rational for agents to prefer the outcomes with higher agent-relative value to outcomes with less. Agent-relative values relate to facts about the agent. For instance, the agent who desires knowledge will assign a high agent-relative value to states involving learning. In the typical case, it’s in an agent’s rational self-interest to pursue states with higher agent-relative value rather than states with lower agent-relative value because it will satisfy the agent’s desires. These values are important for self-interest because of the way they connect to the agent. In the case of first-personal grounds, the existence of these grounds explains why a preference is rational in the same way that agent-relative values do. There’s a fact about the agent—what the agent wills—that explains why the preference is in that agent’s rational self-interest. For instance, it is rational for the agent to prefer to keep his vows because by making the vows, the agent has created considerations in favor of keeping the vows. The preference is appropriate given that fact about the agent.

Perhaps one difference between agent-relative values and first-personal grounds is a temporal one. Agent-relative values are typically known earlier than the decision
whereas first-personal grounds are created simultaneous with the decision. But the relationship between self-interest and its grounds is not a diachronic one—or at least there is no reason to think that it is. Once we look past this temporal difference, we see that first-personal grounds and agent-relative values are analogous in the way they connect facts about the agent to facts about what is in the agent’s self-interest. First-personal grounds are in no worse position than agent-relative value. In fact, first-personal grounds may be less arbitrary than agent-relative values. Agents often cannot control what desires they have. Willing, however, is an expression of agency and flows form the agent as opposed to being forced upon the agent. It also is not arbitrary, or at least is no more arbitrary than the desires an agent may find himself having. If agent-relative values can survive this objection, then it seems that first-personal grounds can as well.

Finally, if we can create first-personal grounds non-arbitrarily, and the range of grounds we can create is unrestricted by value, it seems that we can create first-personal grounds as a way to bootstrap our way to a decision that is otherwise rationally impermissible. For instance, suppose that promising to perform an act does create new grounds that make performing that act rational though doing so without the promise would not have been. It then appears that agents can then promise to do things in order to make them rational. For instance, it would be irrational for me to prefer to adopt a puppy, but if I promised to do so, then I might be able to make it rational for me to have that preference.

One way to block this objection is to allow that people can create reasons willy-nilly, but that it’s only rational to do so in cases of underdetermination. If given reasons determine that it is irrational for me to prefer to adopt a puppy, I can create additional reasons, but it would be irrational to do so. Another way to block this objection is to only allow for the possibility of creating first-personal grounds in cases where reason underdetermines the outcome that an agent ought to prefer.
Chang, for instance, takes this route. Her hybrid version of voluntarism is distinctive from other versions of voluntarism because on it, agents can only create reasons in underdetermination cases. Since the voluntarist reasons only arise in these scenarios, they cannot go against reason when reason is definitive, thus blocking the objection. A response along these lines seems appropriate in the case of first-personal grounds and self-transformation as well. I’m not advocating the use or creation of first-personal grounds in ordinary situations where value determines what one rationally ought to do. Indeed, I’m not convinced that agents can prefer what they judge to not be in their own rational self-interest. If that’s right, then it’s not possible for agents to create first-personal grounds—they would already have the preference. However, in cases where value underdetermines what one rationally ought to do, it seems perfectly appropriate for agents to create their own reasons. Furthermore, the preconditions for creation are restricted to cases of underdetermination by value, so it’s not the case that agents can create reasons to bootstrap their way to a preference that is otherwise rationally impermissible.

3.7 Metaphysical and Ethical Implications

I want to begin this section by considering an analogy between decisions involving self-transformation and decisions involving non-identity. In *Non-Identity*, global values are not enough to determine what would be in the self-interest of an agent. To a disinterested third party, the global values can ground a preference that Wilma had waited to conceive. However, from Wilma’s child’s perspective, global values underdetermine what she ought to prefer since they are not specific to her. From the standpoint of self-interest, the child needs to first find the potential futures in which she exists. Then, of the futures that contain her metaphysical profile, she can look at the agent-relative values associated with her potential future self. Of those, she ought to pick the one in which she enjoys the most value.
In self-transformation cases, this procedure of narrowing potential futures to those containing the relevant metaphysical profile and then looking at agent-relative value fails. Metaphysical profiles and the agent-relative values associated with them are too general to pick out what matters to the agent. Let the futures that contain the agent and preserve the features the agent regards as essential be futures that contain the agent’s *practical profile*. From the standpoint of self-interest, agents need to find potential futures that contain their practical profiles before taking value considerations into account. Then once the potential outcomes have been narrowed to those including the relevant practical profile, value considerations can come back into play. Crucially, value alone doesn’t fully ground rational preferences since it can’t establish the nature of an agent’s practical profile. The above discussion suggests that futures that do not contain an agent’s practical profile are on a par with futures that do not contain an agent’s metaphysical profile—in both cases, opting for the outcome without the relevant profile over an alternative that contains the relevant profile is not in the agent’s self-interest.

This similarity contains lesson for the connection between metaphysics and ethics. Curiously, ethics often takes on board substantive metaphysical theses or is willing to defer to what metaphysics says. Take, for instance, the non-identity problem. In order for the problem to get going, a version of origin essentialism—that the specific sperm-egg combination a person develops from is essential to the metaphysical identity of that person—must be true. But origin essentialism is a controversial metaphysical thesis that many metaphysicians do not accept. Similarly, consider the orthodox decision-making procedures we’ve been discussing in this paper. When we think about agent-relative values of future agents, we take for granted that future agents

---

13 Origin essentialism has its defenders such as Forbes (1985) and Salmon (1981). However, it also has many opponents. In one camp, anti-essentialists like Quine reject origin essentialism as a natural extension of their rejection of essentialism. Even amongst essentialists, the position faces seriously challenges. For instance, Robertson (1998) offers counterexamples that purport to show that a thing can survive slight changes to its origin.
are identical to the present agents under consideration. At best, decision procedures are neutral with respect to the metaphysical questions, and are willing to defer to whatever metaphysics decides is the correct view of personal identity. Let’s call this methodology that resolves metaphysical questions before ethical ones a *metaphysics-first ethics*.

In contrast, consider an alternative methodology that resolves ethical questions before metaphysical ones, an *ethics-first metaphysics*. This ethics-first methodology is less common than the metaphysics-first methodology, perhaps because many take metaphysics to be more fundamental than ethics. I’d like to suggest that this attitude is mistaken, and that significant progress in metaphysics might be made by taking a considered, ethics-first approach.

A modest version of ethics-first metaphysics is put forth by ?. Dasgupta points out that there are many overlapping person-like entities that we can quantify over. There is the entity that origins essentialism gives us, and there is the entity that is picked out by ‘X’s first-born child.’ When we consider which entity is the relevant entity for our concern, taking the metaphysics-first approach might push us towards the entity origins essentialism gives us, perhaps because it is less disjunctive. If we’re concerned about this entity, we find ourselves having to get out of the non-identity problem. But if we take an ethics-first approach, we start with the intuition that it is wrong for the parents to have the less well-off child, and then see that what we care about is the entity that is picked out by ‘X’s first-born child.’ Starting with a piece of ethical data shows us which entity matters practically, and suggests that our metaphysical inquiries should be directed towards that entity. Let this modest version where ethics tells metaphysics which entities are of special practical concern, *practical ethics-first metaphysics*.

A slightly stronger version of ethics-first metaphysics is *epistemic ethics-first metaphysics*. In this version, ethical facts bear evidentially on metaphysical ones.
For instance, in the non-identity problem, the entity picked out by ‘X’s first-born child’ is a disjunctive entity. It could be the child conceived this month, or the child conceived next month, and so on. Dasgupta takes the fact that there is this disjunctive entity of special concern to us to suggest that a particular version of essentialism is true. The judgment that it is wrong for the parents to have the less-well off child leads to an interesting metaphysical thesis, namely ‘subtractionism,’ Dasgupta’s favored version of essentialism.\footnote{For an in depth discussion of epistemic ethics-first metaphysics, see Chan’s “Ethics-First Metaphysics,” or chapter 1.}

If we take this ethics-first approach seriously, then our earlier discussion of the problem of self-transformation may hold some interesting metaphysical implications. I’ll close this paper by discussing what some of these implications might be. First, similar to what Dasgupta points out in his paper, the entities that practically matter are not the ones that are picked out by the leading accounts of personal identity. This point is not particularly novel—Parfit makes it in Reasons and Persons when he famously argues that personal identity is not what matters. Like Dasgupta, I’ve suggested that there is a distinction between a thing’s metaphysical profile and a thing’s practical profile. Metaphysical profiles tell us what a thing is. A thing or person’s metaphysical profile contains all of its metaphysically essential properties, such that what that thing is involves those properties and cannot survive the loss of them. Practical profiles only apply to agents, and they tell us who an agent is. A person’s practical profile contains all of its practically essential properties, such that who that person is as an agent involves those properties and cannot survive the loss of them.

Second, and perhaps a bit more novel, it turns out that for each agent, there are more candidate practical profiles than there are candidate metaphysical profiles. Suppose that an agent’s metaphysical profile is settled. For instance, we might know
all of the young Russian’s metaphysically essential properties and know for any given change, whether the young Russian survives that change. In this scenario, we might imagine that there are still multiple candidate practical profiles. One such profile might include the practically essential property of having socialist ideals. Another candidate profile might not include that property. Holding fixed the metaphysical profile of the agent, we can imagine various practical profiles that might belong to that agent. This result, combined with the solution based on first-personal grounds, leads to an interesting implication. If I’m right about that solution, then it turns out that we can determine which practical profile belongs to us in a rather robust way.

Recall Parfit’s young Russian who regards his socialist ideals as essential to who he is. On almost all accounts of essence, having socialist ideals is metaphysically inessential. The young Russian will continue to exist regardless of if he loses those ideals. But when we try to figure out which of many person-like entities who overlap with young-Russian practically matter, we might find that the entities who maintain the socialist ideal are the ones who matter. But the only reason the having the socialist ideals is essential is because the young Russian has decided that it matters. By willing that certain features matter, agents can pick out which entity is the one that matters and bring it into existence through action. Thus, there is an even stronger version of ethics-first metaphysics at play. Let this version be a determining ethics-first metaphysics.

Here’s a final potential implication. (I say potential because what I’m about to say is controversial and I’m not going to defend it here.) Suppose, as a non-trivial number of metaphysicians do, that the task of metaphysics is to figure out what’s fundamental. When we look at the landscape of person-like entities, part of what metaphysicians are trying to work out is which of these entities is most fundamental. Potentially—and here’s the point I won’t defend—practical entities are amongst these most fundamental entities. If this is the case, then agents, in virtue of acting and
thus picking out which entities matter, can also determine what some of the most fundamental entities are.

3.8 Cited Works


Abstract. Transformative experiences are epistemically and personally transformative: prior to having the experience, agents cannot predict the value of the experience and cannot anticipate how it will change their core values and preferences. Paul (2014, 2015a) argues that these experiences pose a puzzle for standard decision-making procedures because values cannot be assigned to outcomes involving transformative experience. Responding philosophers are quick to point out that decision procedures are built to handle uncertainty, including the uncertainty generated by transformative experience. My paper enters here and contributes two points. First, religious experiences are transformative experiences that are especially resistant to these responses. Second, a procedure that appeals to voluntarist reasons—reasons arising from an act of the will—can allow an agent to rationally decide to undergo or avoid an outcome involving transformative experience. Combining these two points results in some interesting implications with respect to practical aspects of religion.

4.1 The Transformative Experience Puzzle

L. A. Paul (2014, 2015a) argues that transformative experiences present a problem for standard decision-making procedures. Transformative experiences are experiences that are epistemically and personally transformative. An experience that is epistemically transformative is one whose subjective value for an agent is unprojectable prior to that agent’s having the experience. In other words, the subjective value of the experience is inscrutable, or indeterminate, until it has been undergone. Personal transformation adds another dimension of inscrutability.
Personal transformation radically changes what it is like for a person to be him or herself; for instance, the person’s core preferences may drastically change. Parenthood, one of Paul’s token examples, exemplifies both of these sources of inscrutability. Agents who are not parents simply are not situated to know what being a parent will be like for them. Parenthood is radically different from non-parenthood, and the experience of it is one “with an epistemically unique phenomenal character, and moreover, one that can be personally transformative” (Paul, 2015a, 9). Importantly for Paul, the relevant type of knowledge needed in order to assign value is experiential, so from the perspective of an agent who is not yet a parent, the value of parenthood is doubly inscrutable—both the value of the experience and how one will value other things are unknowable prior to the experience.

These two layers of inscrutability present in transformative experiences create a puzzle for standard decision-making procedures. First, as a result of the epistemically transformative aspect, a value cannot be assigned to the transformative experience. Second, as a result of the personally transformative aspect, the way an agent values other things may change, which makes accounting for different future preferences impossible even if one has already worked out how to take known future preferences into account. Thus, if a transformative experience is a potential outcome of a decision, that outcome cannot be assigned a value. Standard decision-making procedures such as ‘maximize expected value’ can proceed only after the values associated with potential outcomes have been ascertained. Even a general procedure that involves weighing reasons in favor of one

1 The personally transformative element won’t pose a problem for decision procedures that completely discount future preferences and take into account only present preferences, but these procedures still face Paul’s puzzle since they must deal with the epistemically transformative aspect. As we’ll see later in the paper in §4.3, personal transformation is challenging in another respect—it forces an agent into a choice that must be made “authentically.” On Paul’s account of what it is to choose authentically, it’s not clear that ignoring future preferences circumvents the challenge of choosing authentically.
outcome against countervailing reasons typically cannot proceed unless key reasons (e.g., pragmatic reasons) are known. Without the ability to assign values for outcomes involving transformative experiences, the decision-making procedure can’t even get started. As a result, the following puzzle arises: how can an agent rationally decide to undergo or avoid a transformative experience on the basis of any standard decision-making procedure?

4.2 Transformative and Religious Experience

4.2.1 Responding to Paul’s Puzzle

As many philosophers are quick to point out, standard decision-making procedures can handle uncertainty. Uncertainty is an essential feature of gambling scenarios, and it is possible use standard procedures in these cases. Similarly, those responding to Paul argue, the uncertainty involved in the paradigmatic cases of transformative experience does not preclude using standard decision procedures.

There are different approaches one might take to avoid Paul’s puzzle. First, one might turn to Paul’s paradigmatic cases of transformative experience such as parenthood and deny that one cannot know anything whatsoever about what the experience will be like prior to having the experience. As long as one has a sufficient grasp of what the experience will be like—perhaps one can have a general idea of what the experience will be like even if one does not know exactly what it will be like—one has enough information to rationally decide to undergo or avoid the experience. If one can know enough about what experiences are like for sufficiently many paradigm cases, then Paul’s puzzle loses much of its force. Meena Krishnamurthy (2015), for example, responds to Paul’s parenthood example by arguing that “we can on the basis of experiences of similar types know what it is like to have a child” (180). One might reflect upon previous experiences involving responsibility over other living
things such as taking care of a cat or babysitting. Elizabeth Harman (2015) makes a similar point. Caring for a younger sibling in a quasi-paternal way and witnessing the relationship between a close friend and her first baby provide at least some evidence of what the experience of parenthood is like.

Nathaniel Sharadin (2015) offers another variation of this type of approach. Sharadin suggests that an agent can reflect upon the non-phenomenal elements of the experience, such as the financial burden and sleep deprivation involved in parenthood, and consider how those things typically affect the agent. The core idea in Sharadin’s account is that there is a supervenience relation between these non-phenomenal elements and an agent’s phenomenal experience. The phenomenal nature of an experience supervenes upon the non-phenomenal elements of that experience, so knowing enough about the non-phenomenal elements of an experience allows an agent to predict what the phenomenal experience will be like. At the very least, an agent will be able to predict the valence of the phenomenal experience.

If Krishnamurthy, Harman, or Sharadin are right, then one can know something about what parenthood will be like. Even though this knowledge is impoverished in comparison to actually having the first-person experience, it is enough to predict what the value of the experience will be like. If we take transformative experiences to be a functional concept that includes paradigmatic cases like parenthood, then this approach shows that one can apply standard decision-making procedures to cases of transformative experience.

There is a second approach one might take to arrive at value. Dougherty et al. (2015) draw a distinction between what an experience is like and how valuable the experience is. Though direct experience provides a route to value, it does not provide a unique route. Testimony from those who have had the relevant experience can provide information about value. In their “Mystery Closet” thought experiment,
1000 customers undergo a rare and previously unknown experience. Though none of the 1000 disclose what the experience is like, they all report that they value the experience. This testimony allows one to predict that the value of the experience will be positive even though one remains in the dark with respect to what the experience is like.

The central theme running through these responses and others is that the value of a transformative experience is not completely inscrutable. These responses apply not only to the parenting case, but to other alleged cases of transformative experience such as tasting Vegemite, becoming a vampire, seeing redness, or committing to a deep personal relationship. Via various types of evidence, we can in fact project what the value of these experiences will be like. This projection can take the form of a broader or narrower range, but as long as the range is bounded—which it presumably is on each of these responses—most decision-making procedures can be used. In special cases, a range that is unbounded in only one direction might also suffice for making a rational decision. For convenience, I’ll give a name to this very general method and call it the ‘Project A Range’ method, or ‘PAR’ for short. Significantly, two conditions must be satisfied for PAR to be applicable: (i) there is some evidential route to projecting the value of the experience, and (ii) the projected value falls into a bounded range (or a range bounded in at least one direction).

\^{2}Richard Pettigrew (2015) offers what is essentially another variation on this theme. He proposes that outcomes be more fine-grained and include value. For example, the agent deciding whether to become a parent should not think of parenthood as a single outcome and then try to assign a value to it. Rather, the agent should consider multiple parenthood outcomes, each with different values. The agent can then assign probabilities to each of the fine-grained outcomes by doing empirical research to see how frequently each outcome occurs. This process is still relies on the potential projectability of value since presumably the point of empirical research is to establish a range of values that correspond to the experience.

\^{3}Cases such as the St. Petersburg paradox show that there are special difficulties involved in making decisions with infinite payouts. But presumably, some decisions involving infinite payouts can be resolved easily (e.g., playing a fair coin toss game where one loses $1 on tails and receives an infinite payout on heads), so a transformative decision involving a range that is unbounded in only one direction might not create a problem.
PAR most obviously applies to the epistemically transformative element of transformative experience. The rationale behind it also applies to the personally transformative element. Just as one can draw from previous similar experiences to form expectations about what a transformative experience will be like, one can draw from those experiences to predict how one’s core preferences might change. In Harman’s case, for example, one reasonably could predict that one’s preferences have shifted from being self-oriented to being child-oriented. Alternatively, one might extend Dougherty, et al’s, point about testimony to the personally transformative element to predict how one’s future preferences might change. As long as one’s post-personal transformation preferences can be projected within a bounded range, one’s preferred diachronic decision-making procedure can be applied. It thus turns out that the inscrutability associated with the epistemic and personal transformative elements of cases like parenthood don’t preclude using standard decision procedures.

At this point, the dialectic between Paul and her respondents can go in two different directions. First, the debate might center on paradigmatic cases of transformative experience, such as the parenting one, and whether these cases are in fact cases of transformative experience. Second, the conversation might turn to the nature of personal transformation and whether the challenge it presents involves more than projecting one’s future preferences. The remainder of this section addresses the former, and the next section returns to the latter.

Part of the resulting debate turns on whether particular cases are in fact transformative experiences. For instance, Paul (2015b) responds to Harman by pointing out that the previous experiences of caring for the younger sibling is of the parenting experience type. Thus, Harman’s point is not a counterexample to Paul’s claim since the knowledge of what the experience of parenting is like is based upon the experience of parenting. In the other direction, respondents apply PAR to
putative examples of transformative experience that Paul provides. For instance, Krishnamurthy suggests that if Jackson’s Mary were exposed to pink in addition to white and black, Mary might have a sufficiently analogous experience (i.e., her experience of pink) and could use that experience to predict what her experience of red might be like. Krishnamurthy then goes on to argue that the parenting case is more like the case of the Mary who’s experienced pink than the Mary who’s only experienced white and black.

I do not propose that we weigh in on these specific cases. Instead, I want to consider what the success of PAR in specific cases would show. For instance, even if the parenting case is not a genuine case of transformative experience, there might still be other cases of transformative experience that give rise to Paul’s puzzle. If there are some possible cases in which an agent faces a transformative experience, then the puzzle stands and we must confront the question of what an agent ought to do in those cases. Of course, we don’t want the only cases where Paul’s puzzle stands to be the contrived inventions of philosophers (though many philosophers may be perfectly happy considering the implications of such cases). Chang makes precisely this point. She is willing to grant that a de novo exercise of a basic capacity (e.g., a new sense modality) where there is no available testimony or information regarding what the experience is like might count as epistemically transformative. But since there is testimony about basic human capacities, such as the five senses, epistemically transformative experiences are “difficult to imagine”, and it would be undesirable for paradigmatic transformative choices to “involve such fanciful experience” as becoming a vampire or gaining the ability to teletransport at

4It is not difficult to imagine possible transformative experiences. For instance, suppose that Nozick’s experience machine can generate literally any possible experience and selects experiences at random; one’s loftiest dream can come true, but one’s worst nightmare can also be realized. An agent deciding whether to be plugged into the machine faces a choice involving a transformative experience. While this case is interesting in its own right, it’s not as pressing as cases, like becoming a parent or undergoing a religious experience, that many agents do in fact face in their every day lives.
will (249). The puzzle is more pressing when it involves a transformative experience that is a live option for many people. I propose that there is at least one such case—the case of religious experience.

### 4.2.2 Religious Experience

Religious experience presents an interesting case of transformative experience. I take religious experiences to be experiences that the subject takes to be an experience of some sort of ultimate or divine reality and have religious import. These experiences may or may not be veridical; I'll speak of them as if they are, but as I'll explain later, whether these experiences are in fact veridical does not matter for our purposes. Religious experiences also are varied. Some, perhaps such as watching a beautiful sunset and feeling divine presence, are more mundane, while others, such as the beatific vision, are much more pronounced. It is these latter types of religious experiences that are candidates for transformative experiences and are the ones that I'm concerned with in this section. Though occurrences of these experiences are not widespread, they are ones that many people aspire to have and take steps towards having. Thus, these religious experiences present cases of decisions that are live for many people. If these experiences turn out to be transformative, then we will have found a case where Paul’s puzzle arises for a significant number of people.

It is *prima facie* plausible that some cases of religious experience—perhaps at least the more extreme ones—are transformative. Paradigmatic cases of transformative religious experiences might include the beatific vision, conversion, and mystical experiences. People who undergo them tend to find that they could not have predicted what the experience would be like. What the religious experience is like is so radically different from ordinary experiences that they have difficulty explaining them to people who have not undergone them. This ineffability
is characteristic of most religious experiences, which is perhaps what prompts William James to suggest that ineffability is a mark of religious experience. Job, for example, is overwhelmed when he encounters the whirlwind and then finds that he cannot make his comforters understand what he has witnessed. Furthermore, people who undergo religious experiences often find that their worldviews and core preferences drastically change as a result of the experience. The apostle Paul’s conversion on the road to Damascus exemplifies the radical personal transformation that may occur as a result of a religious experience. Religious experiences can be epistemically and personally transformative, and thus qualify as transformative experiences.

Some cases of religious experiences are involuntary—Paul the apostle, for instance, did not choose to be blinded by God’s light. In order to generate the puzzle for decision theory, we need cases that involve a voluntary decision on the part of an agent. Fortunately, there are many such cases. People reflect on whether to subscribe or convert to various religions. Mystics engage in practices such as meditation designed to bring about the religious experience. Perhaps more commonly, people attend religious services seeking some sort of union with the divine; many hope to eventually obtain the beatific vision. In more extreme cases, some people turn to drugs to facilitate a connection with the divine or transcendent. Additionally, many of these cases involve agents who pursue the experience while acknowledging that they do not know what the experience will be like and that their lives may be changed as a result of the experience. Prima facie, these cases look precisely like the type of cases that would generate Paul’s puzzle.

More importantly, the case of religious experience poses a dilemma for PAR that makes the general strategy inapplicable. Recall that in order for the strategy to work, there must be an evidential route that allows an agent to project a range that is bounded in at least one direction. Via the law of excluded middle, either it’s not
the case that there is such a route, or there is. On the first disjunct, which I’ll call the first horn, the first requirement for PAR—that there be an evidential route that enables projecting a range of value—is not met. On the second disjunct, which I’ll call the second horn, there is relevant evidence. However, the motivation behind taking this second horn, I will argue, results in a projected range that is unbounded in both directions, thus failing the second condition for PAR.

The first horn, on which there is no evidence that enables projecting a range of value for religious experience, straightforwardly renders PAR inapplicable. But rather than immediately moving on to the second horn, it will be helpful for the rest of the discussion to take a moment to consider why one might be tempted to take this first horn. Standard cases of transformative experience like parenthood involve a finite object of experience. These objects have properties with which human agents can have acquaintance via either description or acquaintance. These properties, which are shared by other objects with which agents may be familiar, are the grounds of the possibility of finding analogous cases. The more similar two objects are with respect to properties, the more similar experiences of those objects are. The more similar experiences of those objects are, the more similar the value attributed to those experiences will be. This correlation between similarity of properties and similarity of experiences is what makes PAR so promising.

Unlike the standard cases, religious experience involves a connection to a transcendent, and perhaps infinite, object. Because the object is transcendent, predicates, which express ordinary properties with which human agents are familiar, cannot be applied to the object. This is the case even when the object is revealed to the agent. Aquinas (ST I 7.12) argues that human beings cannot intellectually “see” the essence of God on their own, much less fully comprehend God. Even with the

5ST’ stands for Summa Theologiae and is accompanied by the part, question, and article, and quotes are from Aquinas (2006).
assistance of divine illumination that leads to some sight, human beings still will not be able to comprehend God due to God’s infiniteness. This situation involving transcendence is worse than that of the proposed cases of transformative experience thus far. Mary in the black and white room can at least know facts about color properties even if she doesn’t have direct experience of them. A completely blind Mary would at least know something about the role that sense modalities play since she understands what her other senses give her. Furthermore, once Mary gains the relevant experience, she will gain familiarity with color or sight facts. But one cannot be attain the same degree of familiarity one has with ordinary properties when it comes to properties of the divine, even once one has experienced them. Thus, there won’t be evidence from which one can draw to predict what the experience of the object will be like. Taking the first horn is not viable.

If one opts for the second horn instead of the first, one supposes that there might be some evidence. One natural thought is that we might draw from the testimony of those who have had religious experiences. But testimony regarding religious experience is problematic. One might think that those who have had these experiences can describe the experience even if the object of the experience remains mysterious. On the contrary, those who claim to have had religious experiences often report that they cannot adequately describe their experience; the experience is ineffable. Aquinas’s diagnosis of the situation might partially explain this ineffability. This situation resembles Dougherty, et al’s, mystery closet. Perhaps if there is uniform, or nearly uniform, testimony from those who have undergone religious experiences, one might be able to project a range of values despite being in the dark with respect to what the experience is like. However, there are two ways in which the religious experience case is different from the mystery closet. First, there isn’t sufficient uniformity in the valence of values assigned to religious experience. (I’ll elaborate on to this point later in this section.) Second, the range of values
assigned to the experience is unbounded. Combining these two points results in a potential range of value that is unbounded in both directions.

Potential non-testimonial evidence also leads to a potential range that runs afoul of the second clause of PAR. Perhaps there are some properties—e.g., goodness—that apply to God analogically, or in a similar yet distinct way than the one in which they apply to humans. An agent could then reflect on how that agent has valued previous experiences with good people, and then use that as grounds to form a judgment regarding the value of an experience with a good God. In ordinary cases, an agent reflecting on her experiences with good people will realize that when other factors are controlled, there is a correlation between the degree of goodness of the objects of the experience and the value of the experience. The better the people who are, the better the experience tends to be. Since God’s goodness is presumably infinitely greater than our own, it would not be unreasonable to suppose that the value of experiencing God’s goodness would be infinitely positive. At the very least, since there is no limit to God’s goodness, there is no reason to believe that there is a limit to the value of a religious experience. To be clear, I am not suggesting that the experience itself will be infinite or that one needs to commit to something controversial like infinite value. Human subjects are finite even if their objects of experiences are not, so it’s hard to see how an experience itself would be infinite. However, it does seem that the value of an experience of perfect goodness might not be limited to a finite range. There are infinitely many finite numbers, so the range of potential values will still be infinite or unbounded even if the potential values are finite.

It may look like PAR can be used on this second horn since the range of value can be unbounded in one direction. If the value of religious experience is unbounded in the positive direction, then a decision involving religious experience begins to look rather Pascalian. There are at least some scenarios under which an agent can rationally
opt for the outcome with unlimited positive value, and this situation starts to look like one of those. Unfortunately, the case is complicated by the fact that people can have wildly divergent religious experiences. The devil’s goes rather poorly for him, and he falls because he is such that he cannot stand in union with God. One might be tempted to write off this story as an atypical case, but if Anselm is correct in De Casu Diaboli, the devil falls because of his “will for happiness,” not because there is something obviously defective with him. The point here is that the goodness of the object of experience is not sufficient to guarantee a valuable experience for the subject; facts about the subject factor into subjective value as well. What Anselm’s story can be taken to show is that human agents, who are situated similarly to Lucifer and are vulnerable to the same desires for happiness, ought not be too confident that they would avoid a similarly disastrous experience. Moreover, unless one is a universalist about salvation, there are theological reasons to hold that some people who experience the divine do not have experiences with positive value. On some accounts of hell, those who are in hell are there because they simply cannot stand being in the presence of God; the beatific vision has negative subjective value to them. It is certainly possible that one might have a negative response to a religious experience, especially given the types of creatures humans are, and widely held theological doctrines may even entail this possibility.

It is a mistake to assume that the value of the experience is unbounded in only the positive direction. Clearly, the value of religious experiences can be negative. Combining that possibility with the considerations that resulted in the range of

---

6 Actually, humans are in a far worse situation than Anselm’s Satan because unlike Satan, their choice is made in an imperfect universe under epistemic uncertainty with respect to God’s existence and goodness.

7 One might distinguish between universalists who believe that hell is “locked from the inside” and is populated by people who could in principle leave, and universalists who believe that everyone is eventually brought into union with God. Only the latter might be able to escape this point about a potentially negative valence.
values being positive and unbounded means that the potential values could also be negative—and unbounded. The ultimate result of taking the second horn is that PAR cannot be used since the range of potential values is unbounded in both directions. Such a range forces an agent who is deciding whether to undergo a religious experience back into the original puzzle Paul poses for transformative experience.

Perhaps one might object along the lines of Pettigrew (2015). The outcome involving the transformative religious experience can be partitioned into an outcome with positive unbounded value and an outcome with negative unbounded value. Given that the outcome where the transformative experience occurs obtains, the probability that it is the outcome with positive unbounded value might be substantially higher than the probability that it is the outcome with negative unbounded value, or vice versa. If the probability assignments to these two outcomes are asymmetric—perhaps the positive outcome is twice as likely as the negative one—then perhaps an agent is rational in deciding to undergo the transformative experience.

There are at least three difficulties with this objection. First, decision procedures such as ‘maximize expected utility’ have difficulties dealing with outcomes involving infinite value. Though there is a substantial literature on how to modify decision procedures in order to accommodate these cases, there isn’t a settled resolution and what one ought to do in the case described above is unsettled. Second, in the case of transformative religious experience, it’s not clear what the grounds on which probabilities could be assigned to the partitions are, or even whether there are such grounds. But, third, even if there were grounds for assigning probabilities (perhaps exhaustive empirical studies tell us that two-thirds

\[8\text{This issue frequently arises in the context of Pascal’s Wager. See, e.g., Bartha (2007); Hájek (2003).}\]
of the people who undergo transformative religious experience assign it a positive value), there is still the problem of authentically choosing in light of the personal transformation that occurs. This point highlights yet another significant difference between religious transformative experience and non-religious transformative experience: the degree of personal transformation that occurs. Standardly, faith is thought to transform people totally; every aspect of the agent’s life is affected by the adoption of the new perspective. The pervasiveness of transformation in the agent’s life in the case of religious experience might be contrasted with the other paradigmatic cases in which personal transformation might occur to a lesser degree.

I’ve shown that religious experiences are a type of transformative experience that resists PAR. Furthermore, religious experiences are transformative experiences that people regard as live outcomes in the choices they make. Perhaps one might try to resist the case of religious experience by arguing either that religious experiences of the sort I’ve described are impossible or that infinite beings are impossible. I suspect that the former collapses into the latter. One might think that religious experiences of the sort I’ve described are impossible because there are no divine beings or because divine beings cannot be infinite. While there are controversial arguments that allegedly demonstrate that a perfect being is impossible, there are no arguments that show that non-perfect divine beings are logically impossible.

As long as the chances of having a religious experience of the sort I’ve described are non-zero and non-infinitesimal, standard decision-making procedures will encounter difficulties. Consider playing a game of roulette, where you will win a religious experience if the ball lands settles on double zero. Even though the chance of winning the religious experience is only 1/38, the value of the religious experience enters the procedure. Since the value ranges over all the negative and positive numbers, maximizing expected utility, using dominance reasoning, and other strategies will not work. The same holds true in mixed strategies. For example, if one were to
flip a coin in order to determine whether to play the roulette game, the inscrutable value would still enter into the decision-making procedure. Thus, the case of religious experience is a special case of transformative experience that enters the picture even if the possibility of religious experience is uncertain.

4.3 Voluntarist Reasons and Transformative Experience

In this section, I want to shift gears and consider whether there is a rational way of deciding to undergo or avoid a transformative experience of the sort that Paul envisions. Paul has her own preferred solution involving revelation, but I’d like to take a different tact. I’ll focus on the role of voluntarist reasons, which are reasons created by agents through an act of willing. To be sure, whether there are such things as voluntarist reasons is controversial. My purpose in this paper is not to enter this debate; instead, I’ll suppose that there are voluntarist reasons and explore how they might offer a solution to Paul’s puzzle. As I’ll explain, voluntarist reasons are often used to resolve decisions in cases of parity. These cases of parity have certain features that make using voluntarist reasons appropriate. Cases of decisions involving transformative experience have these same features, and I’ll argue that these features make using voluntarist reasons appropriate in transformative experience cases as well. I’ll end by circling back to Paul’s preferred solution and discussing how my solution connects with hers.

4.3.1 Voluntarist Reasons

Ruth Chang introduces voluntarist reasons to resolve cases of parity, or cases in which the relations of better than, worse than, and equal to do not hold between the potential outcomes. In these cases, one’s objective reasons—e.g., pragmatic or moral

---

9Chang (2009, 2013) and Korsgaard (1996) are the two most notable defenders of voluntarist reasons. For arguments against see, e.g., FitzPatrick (2005); Railton (2004); Scanlon (2010).
reasons—underdetermine what one ought to do. For example, suppose an agent is deciding between pursuing a career in philosophy or a career as a professional athlete. The agent might run through all of her objective reasons for preferring one career over the other. For example, pragmatic reasons might favor a career in philosophy because the agent values having an exotic, flashy career, while moral reasons might favor a career as a professional athlete since the agent would have the larger public platform from which social causes could be advanced. In addition to pragmatic and moral reasons, the agent might take into account other objective reasons such as epistemic, legal, and aesthetic ones. After summing all of the reasons in favor of each career path, the agent might discover that the objective reasons underdetermine the choice. Neither choice is better than the other (otherwise the decision would be clear) and the choices are not equally good. If the careers were equally good, adding a slight perk to one of the careers, such as increasing the annual salary in the philosophy career by five dollars, would tip the scales in favor of that career. But this “small addition” doesn’t tip the scales, which means that the careers are not equal. Rather, they are on a par with each other.

Cases of parity, like cases of transformative experience, also present difficulties for standard decision-making principles. Chang’s solution to cases of parity relies on what she calls voluntarist reasons. Unlike objective reasons, voluntarist reasons are created through an act of the will of the agent. They come into play when objective reasons, such as epistemic or pragmatic reasons, underdetermine the choice that an agent is facing. In such cases, agents can give more weight to a reason by willing that the reason matter more. For instance, our agent deciding between becoming a

---

10 Whether the types of reasons I’ve listed represent distinct categories of reasons, reduce to each other, or are species of practical reasons does not affect the central point, which is that reasons can underdetermine the outcome for which one ought to opt. I list these various types of reasons to illustrate the considerations that rational agents might take into account, and am agnostic about how they relate to each other and their ontological status.

11 For Chang, voluntarist reasons strengthen existing objective reasons. Other defenders of
philosopher or athlete might will that the pragmatic reasons matter more to her. By
doing so, she gives more weight to that reason. This strengthened reason can then
tip the scales in favor of the philosophy career. Most importantly, in willing that a
reason matter more, the agent reveals herself to be the type of person for whom the
decision is appropriate, and thus can rationally opt for that decision.

Finally, voluntarist reasons cannot tip the scales in cases where objective
reasons are decisive. For example, if our agent were faced with the choice of being a
professional assassin or being a philosopher, the objective reasons would strongly
favor being a philosopher; voluntarist reasons would not come into play, and would
not tip the scales even if they did. Because the “correct” results are maintained in
cases where objective reasons determine the decision, voluntarist reasons are
compatible with one common sense way of making decisions—weighing objective
reasons and then seeing to which choice those reasons lead. But voluntarist reasons
add the ability to rationally make a decision in cases where objective reasons
underdetermine which outcome should be chosen.

4.3.2 Applying Voluntarist Reasons to Transformative Experience

With respect to transformative experience, Chang draws a distinction between
event-based and choice-based transformative choice. An event-based transformative
choice is one where “an event or process downstream from a choice—perhaps
experiential—transforms you”; a choice-based transformative choice is one where
“the making of the choice itself transforms you,” (Chang, 2015, 240). Chang argues
that no problem for decision theory arises for event-based transformative choice for
many of the same considerations discussed in § 4.2. However, she suggests that
voluntarist reasons, such as Korsgaard, offer accounts on which voluntarist reasons create new
reasons _ex nihilo_. For the purposes of my paper, I’ll use Chang’s account, but I don’t have a stance
on what the correct account of voluntarist reasons is and both accounts provide the tools to resolve
cases of parity and cases involving transformative experience. See Chang (2009, 2013); Korsgaard
(1996).
choice-based transformative choice does pose a problem and that using voluntarist reasons when facing them is appropriate. While I largely agree with her point regarding choice-based transformation, I disagree with her diagnosis of event-based transformative choices. Furthermore, these choices are extensionally equivalent to the choices involving transformative experience that Paul raises. I'll thus address what Chang says about event-based transformative choice and argue that voluntarist reasons apply to those choices as well.

Event-based transformation occurs when an event downstream from a choice is transformative. For example, the experience of becoming a parent (assuming that it is in fact transformative) is an event-based transformation. Learning what it is like to become a parent occurs after the choice to become a parent, and insofar as that experience is epistemically and personally transformative, it is an event-based transformation. As mentioned in §4.2, Chang believes that we do not in fact encounter these sorts of event-based transformative choices. The proposed cases of parenthood or gaining a new sense modality do not qualify because we can know enough about them to have some sense of what those experiences will be like and what their values will be. But as I argued, the case of religious experience is one that is both transformative and a pressing issue for many people. Thus, there is at least one event-based transformative choice that people face, and the question of whether one can make a rational choice in these cases arises.

Considerations in favor of applying voluntarist reasons to cases of parity also favor applying them to the case of religious experience (and any other cases of transformative experience). Voluntarist reasons come into play in parity cases since objective reasons underdetermine the choice the agent faces. Something similar

---

12For Chang, choice-based transformative choices do not need to involve transformative experiences. The examples she gives of these choice-based transformative choices are essentially ones in which there is parity between outcomes. These choices are “transformative” in the sense that they change an agent’s normative reasons.
occurs in cases where the value of a potential outcome is inscrutable. Let’s return to the case of the agent deciding whether to become a parent (and grant that the value of being a parent is in fact inscrutable from that agent’s perspective). Unlike normal cases of parity, it could turn out that becoming a parent turns out to be significantly better than being childless, or vice versa. However, the fact that either of these could be the case from the agent’s epistemic position creates the same underdetermination by objective reasons that is present in parity cases. The agent doesn’t know which outcome is favored by pragmatic reasons if the value is unknown. Furthermore, there don’t seem to be other objective reasons that strongly favor one outcome over the other. In most cases there are not overwhelming moral reasons to have a child or to remain childless. Also as in the parity case, the outcomes are on par as opposed to exactly equal. Adding a small monetary benefit such as $50 to becoming a parent does not tip the scales in favor of that outcome. Finally, this is a situation in which the creation of a voluntarist reason in favor of one of the outcomes would demonstrate that the outcome opted for is fitting for the type of person the agent is. In light of these similarities to cases of parity, applying voluntarist reasons to decisions involving transformative experiences is similarly appropriate.

Let’s consider how a decision procedure involving voluntarist reasons would be applied to a decision involving transformative experience. Suppose that the agent considering whether to become a parent recognizes that objective reasons underdetermine the decision. But this agent wills that she cares a lot about genetic reasons, and these push her towards having a child. By willing that these reasons matter more, the agent strengthens that reason in favor of having a child. The additional weight of that reason now might tip the scales in favor of having a child, despite the fact that the value of having a child is inscrutable. Furthermore, by willing that genetic reasons matter, the agent reveals herself to be the sort of person
for whom the choice is rational. If instead she had willed that avoiding a potential transformation mattered more (perhaps she suffers from status quo bias), then she would have revealed herself to be the sort of person for whom it was rational to avoid the experience. Thus, the willing of a reason to matter more provides a way for the agent to rationally arrive at a decision in cases involving inscrutable values, such as transformative experiences.

4.3.3 Two Objections

This application of voluntarist reasons may seem controversial, and there are two potential objections one may have. The first is one that Paul (2015b) makes and applies specifically to whether it is appropriate to use voluntarist reasons in cases involving inscrutability. Paul argues that if an agent faces an epistemically and personally transformative outcome, the agent cannot grasp the subjective value of that outcome. Without grasping the subjective value of an outcome, the value is “noncomparable” in Chang’s sense of noncomparability. But this diagnosis cannot be correct. Chang’s sense of noncomparability has to do with category mistakes with respect to the dimension of comparison. For instance, oranges and the number nine are noncomparable with respect to tastiness, for the number nine is not the sort of thing to which tastiness applies.\footnote{For more on Chang’s account of noncomparability, see Chang (2014).} Clearly, no such category mistake is being made in the case of subjective values. Even if the values are inscrutable in the sense of being unknown or ungraspable, the comparison at hand is a comparison between two subjective values of an experience. These two subjective values are of the same type, and are thus not noncomparable.

Second, one might object to focusing on reasons for bringing about an outcome rather than the value of the outcome in the context of decision making. First, reasons tend to co-vary with value, and thus won’t yield radically different results.
For example, if an outcome has a value of ten hedons, there is a corresponding pragmatic reason in favor of bringing about that outcome. Focusing on reasons thus takes into account the value of outcomes. Second, on some accounts, reasons, more so than values, are what provide normative force. The amount of value a thing has is a descriptive fact about that thing, and descriptive facts by themselves do not explain why a thing ought to be pursued. Reasons, which take into account these descriptive facts, do carry normative force. Since the transformative experience puzzle focuses on what a rational agent should do, it makes sense to focus on the bearers of normative force. Thus, considering the various reasons in favor of undergoing or avoiding transformative experiences is appropriate, and allowing for voluntarist reasons provides a way out of Paul’s puzzle.

4.3.4 Paul’s Solution

How does my proposal compare with the one Paul lays out in Transformative Experience? For Paul, a crucial part of making decisions involving transformative experience is that they be made authentically, or done from a “first personal perspective” rather than a third-person one. An agent must consider what an experience will be like for her. An objective third-person perspective simply is not adequate when the choice before an agent is one concerning what her future will be like for her and who she will become; third-person accounts fail to capture these important dimensions of the outcome. Because rational decision-making is typically associated with the third-person perspective, the problem of choosing rationally and authentically appears insurmountable. This way of understanding the problem brings us back to the second point of contention between Paul and her respondents. There is more to the personally transformative element than the preferences an agent will have post-transformative experience. Being able to project future

---

14See especially Chapter 4: “The Shock of the New.”
preferences may be sufficient for choosing rationally, but it still does not help with choosing authentically.

Fortunately, according to Paul, an agent can choose both rationally and authentically by reframing the decision as one involving revelation, or the discovery of what the agent’s life will be like and who the agent will be, post-experience. Once the problem is reframed in this way, an agent may choose revelation for its own sake. In doing so, the agent has chosen rationally (by choosing in accordance with the preference for discovery) and authentically (by choosing based on wanting to know what the experience will be like for her and who she will become as a result of it).

Choosing based on the desire for revelation is compatible with using voluntarist reasons to choose. The reason one might will to matter more could be a revelation-regarding reason. Furthermore, the decision is rational (because one does what there is most reason to do) and authentic (because one’s willing is done from a first-person rather than third-person perspective). Perhaps Paul’s solution is preferable because it does not take on board any commitment to voluntarist reasons. However, unless one has independent reason to reject voluntarist reasons, there are three advantages to choosing based on voluntarist reasons. First, consider choosing revelation for its own sake as opposed to choosing revelation via a voluntarist reason. On the former, there is no further explanation for why revelation has been chosen for its own sake as opposed to something else besides revelation having been chosen for its own sake. But on the latter, there is an available explanation: the agent has chosen to be the kind of person who chooses revelation for its own sake.

The second advantage is related to the first. In cases of forced choices between two transformative experiences, choosing only on the basis of desire for revelation is arbitrary. Return to the case involving the choice between two careers. On what grounds could one choose the discovery of what it is like to be a professional
philosopher over the discovery of what it is like to be a professional athlete? Revelation for its own sake does not provide a difference maker between the two options since both involve discovery. To avoid arbitrariness, one would have to choose to discover a particular career. But an agent cannot pick in this way based on first personal considerations when she is completely impoverished with respect to what the experience will be like for her. The only non-arbitrary grounds for choice would have to be some objective aspect associated with the particular careers. Using voluntarist reasons is needed to make a non-arbitrary, rational, authentic choice.

Finally, in the case of transformative experiences involving other people, there is something unsatisfying about choosing on the basis of revelation. Take, for example, the case of an agent who is deciding whether to get married. Presumably, the essence of marriage is a certain type of concern for or commitment to one’s partner. But choosing based on a preference to discover what marriage to that person is like does not quite capture concern for that particular person. Something similar might be said in the case of parenthood. Consider an agent who is considering whether to adopt a child. There’s something objectionable about that agent choosing to adopt because he wants to discover what parenthood is like and how it changes him as opposed to choosing for the sake of the child. Voluntarist reasons allow the agent to choose because the agent values the future experiences yet to be discovered with that partner or child, as opposed to the experience yet to be discovered itself. While revelation can allow an agent to choose rationally and authentically, it is not an appropriate way to choose in all cases of transformative experience.

4.4 Voluntarist Reasons and Religious Experience

The points presented in each of the two previous sections are independent of each other; one can accept what I say about voluntarist reasons without accepting what I
say about religious experience, and vice versa. However, if what I say in the both of those sections is correct, some very interesting implications for religious experience arise. Suppose that religious experiences are transformative, and that voluntarist reasons can be applied to decisions with outcomes that involve religious experiences. Two consequences result. First, it can be rational for an agent to pursue a relationship with God even if the agent is uncertain of God’s existence. For example, an agnostic might visit a monastery with the hope of initiating a relationship with God even if the agnostic doubts that the relationship will occur because he doubts that God exists. But as long as the agnostic assigns a non-zero probability to being able to initiate the relationship with God, the value assigned to visiting the monastery is inscrutable when the agnostic decides whether or not to make the trip. Thus, even where an agent is uncertain of God’s existence, it can still be rational for that agent to pursue a relationship with God.

Second, it can be rational to reject a relationship with God. This implication may seem peculiar, but allowing for an agent to rationally reject a relationship with God might actually be a virtue. Take the devil, who has purportedly done just that. If it were grossly irrational to decline to be in a relationship with God, then the devil merely suffers from irrationality. But if the decision involves inscrutable outcomes and an agent can use voluntarist reasons to pick between the outcomes, the will comes into play and the devil is culpable for the decision to spurn a relationship with God. The upshot in the opposite case of the agent who rationally wills a relationship with God is that the agent has demonstrated a praiseworthy decision, as opposed to one that is simply determined by objective reasons.

As far as I know, no one has argued for this second consequence, but I find it appealing and it may shed some light on existing problems in philosophy of religion. For example, this picture might hint at a new model for faith. Faith might just be the willing by an agent to strengthen reasons in favor of seeking a transformative
relationship with God. This willing by the agent is the act of faith. The picture might also provide a partial response to the problem of hiddenness. First, it downplays the claim that those who are not in a relationship with God are worse off. If the value of the relationship is inscrutable, we don’t know whether people would be better off if they were in it. Furthermore, the importance of knowing that God exists in order to pursue a relationship with him is also downplayed since the pursuit can be made rationally even in the face of epistemic uncertainty.

Finally, the picture I’ve sketched has some implications for Pascal’s Wager. On Pascal’s version of the wager, the payoff of belief in God is infinitely positive. However, in light of what I say, the payoff of belief in God is inscrutable rather than infinite. Depending on one’s voluntarist reasons, it can be rational to try to believe in God and enter into that relationship, though with different voluntarist reasons it will be rational not to try to believe in God. This result is weaker than the one intended by Pascal, but if one has the more modest goal of showing that belief or attempted belief is rational, then showing that attempting to believe can be rational accomplishes that goal. There is much more to be explored in the intersection of religious experience and voluntarist reasons, but hopefully this preliminary discussion shows that these issues are worth pursuing.

4.5 Cited Works


