Normative Reality: Reasons Fundamentalism, Irreducibility, and Metaethical Noncentralism

Nicholas Engel

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

Normative Reality: Reasons Fundamentalism, Irreducibility, and Metaethical Noncentralism

Nicholas Engel

Reasons fundamentalists assert that normative reality is constituted by exemplifications of the normative reasons relation: an irreducible, sui generis relation that strongly supervenes on non-normative reality. In this dissertation, I argue that reasons fundamentalists cannot explain why exemplifications of the normative reasons relation strongly supervene on non-normative reality. Irreduciblists about normativity can avoid this problem by asserting, contra the reasons fundamentalist, that normative reality is constituted by exemplifications of thick properties, which provide material for a conceptual analysis of normative reasons. The theory that results analyzes normative reasons for action as answers to questions why an action promotes a thick property.

Nearly every normative theorist affirms what I call

> Additive Normative Supervenience (ANS): Normatively discernible worlds must be non-normatively discernible.

ANS asserts that, if Edward Snowden is morally good, then Snowden's counterparts in worlds that are indiscernible in all non-normative respects must be good. Reasons fundamentalists struggle to explain why ANS is true. I consider and reject potential explanations of ANS that appeal to conceptual entailment and a posteriori necessity. Rosen has recently offered an argument against ANS. Rejecting ANS, however, problematizes irreduciblist accounts of normative explanation and normative epistemology.

Irreduciblists can avoid this dilemma by arguing that ANS is either incoherent or false and adopting an alternative formulation of normative supervenience. Bilgrami's arguments against the intelligibility of normative supervenience doctrines purport to show that ANS is in fact unintelligible, and Merricks' arguments against the supervenience of consciousness on microphysical properties can be extended to show that ANS is false. Neither argument, however, establishes the falsity or unintelligibility of a modified for- mulation of normative supervenience,

> Transformative Normative Supervenience (TNS): Normatively discernible worlds must be descriptively discernible,

where descriptive discernibility is just discernibility with respect to non-normative properties or thick normative properties. Irreduciblists can explain the truth of TNS by adopting non-centralism about normative reasons–that is to say, by maintaining, contra the reasons fundamentalist, that normative reality is constituted most fundamentally by exemplifications of thick properties. This allows the irreduciblist to provide an account of normative explanation and normative epistemology, analyze normative reasons in terms of thick properties, and preserve buck-passing accounts of thin normative properties.

Scanlon has argued that the reasons relation is a four-place relation, relating the facts that are reasons for an agent to perform an action in a given circumstance. I argue that facts are also reasons for an action with respect to a thick property that that action will promote, in contrast to sets of distinct actions that the agent could perform instead. The resulting six-place relation turns out to be an instance of the relation that holds between why-questions and answers. What it is to be a normative reason for an agent to do something is to be a

correct answer to a question why that agent's doing that action will promote a thick property.

Decades ago, Anscombe had also suggested that reasons were answers to why-questions of a certain kind. The attractiveness of this position has been relatively underappreciated in the philosophy of normative reasons, in part because Anscombe had offered the reasons-as-answers thesis as a thesis about motivating reasons rather than normative reasons. The reasons-as-answers thesis also provides resources for those irreduciblists about reasons who reject my non-centralist conclusions to avoid the wrong kind of reason problem for buck-passing accounts of normativity: they can distinguish between right and wrong kinds of reasons by distinguishing between answers to distinct kinds of why-questions.

Contents

Acknowledgements

I wrote most of the material for this dissertation from the spring of 2013 through the spring of 2016. I am grateful to Columbia University for providing dissertation and teaching fellowships for this period, and am also grateful to Katja Vogt, Maia Bernstein, Stacey Quartaro, and Wolfgang Mann for their help to me in securing fellowships for the 2014/2015 and 2015/2016 academic years.

I thank the Columbia Graduate Philosophy Workshop, the Columbia Philosophy Thesis Preparation seminar, the Journal of Value Inquiry conference series, the Alliance program, and the Fordham Graduate Conference for opportunities to receive helpful feedback on earlier presentations of this work.

I thank Carol Rovane, Akeel Bilgrami and Achille Varzi for their help throughout the writing of this dissertation, and to Carol Rovane in particular for sponsoring my doctoral candidacy.

I thank my Hope College professors Carol Simon, Jack Mulder, Jim Allis, Joe LaPorte, and Nick Perovich, for cultivating my interest in philosophy and for encouraging me to pursue a philosophy PhD.

Particular people who have provided helpful comments on early work include Alison Fernandes, Avery Archer, Borhane Blili Hamelin, Chris Whalin, Christine Susienka, Derek

Parfit, John Morrison, Justin Clark-Doane, Katie Macintyre, Katja Vogt, Matt Moss, Max Hayward, Nathan Bice, Nemira Gasiunas, Rush Stewart, Ryan Kemp, and Yang Liu. I thank my many students at Columbia University and The Cooper Union for entertaining my occasional requests to think through my work with them. I thank Alan Klima, Whitney Kazemipour, Lisbeth Fuisz, and all the attention artists who have provided helpful writing momentum.

Personal thanks are due to Francis Hittinger, Grace Delmolino, Isabella Florin, John and Charles Yost, Laura Przybysz, Mary Engel, Michael Klein, Nathan Bice, Robert Moore, Samuel Sawyer and Tommy Hawkins for helping me think through the many financial and social obstacles involved in being a young person and a graduate student in the humanities after the 2008 financial crisis. I thank the residents of Ford Hall, my JVC community, and my Mancation brethren for social support. I thank Lisa Williams for helping me organize my life during the very busy spring 2014 semester. I thank parents for their perpetual support.

Finally, I would like to thank all the normative reasons for being there to non-causally motivate me when I didn't desire to write about them. I would also like to thank them because, while persons have characteristically gotten upset when I've become skeptical of their existence, the normative reasons never did.

For Susanna and her mother

Chapter 1

Normative Reasons

1.1 Introducing reasons

The many thoughts we have about the world (broadly construed) fall into two categories.[1] A first category of thoughts concern the way the world is, what its components are, and how all these components are structured. These are thoughts about the world's descriptive features. A second category of thoughts concern the way the world ought to be and how these oughts are structured. These are thoughts about the world's normative features.

The normative aspect of our thoughts about the world is a rich and complicated landscape. We have deontological notions like those of oughts, norms, rules, rights, and duties. We have thinner evaluative notions like those of goodness, value, and pleasure and pain. We have thicker evaluative notions like "courage" (Aristotle, NE II.2) and "lewdness" (Blackburn and Gibbard 1992). Some concepts are concepts of properties that a thing can have more or less of (goodness is an example), and some, like our notion of rights, are of properties that we have discretely. We have judgment-sensitive attitudes like belief and admiration, and other mental states and emotions that are, as Gibbard likes to say, tinged with a normative character.

[1] There are, of course, many other ways to categorize our thoughts than the categorization I am proposing here.

We also have predicative, attributive, and functional notions of goodness. Geach distinguishes between predicative adjectives and attributive uses of adjectives (Geach 1956). A adjective, used predicatively, is like the adjective "red" in the following statement: "That is a red barn". To say that "red", in this statement, is being used predicatively, for Geach, is to say that the statement "That is a red barn" is semantically equivalent to the statement "That is red and that is a barn". An adjective, used attributively, is like the adjective "big" in the following statement: "That is a big mouse". To say that "big", in this statement, is being used attributively, for Geach, is to say that the statement "That is a big mouse" is *not* semantically equivalent to the statement "That is big and that is a mouse". "Big", in this example, is indeed being used attributively, since it's a mistake to infer from the fact that an object is a big mouse that that object is big simpliciter. "Good" is characteristically being used functionally when it appears in phrases of the form "good for X". When we say that nitrates are good for plants, we mean that nitrates bear some positive causal or constitutive relation to the functioning of plants.[2]

Some of these normative notions, like the distinctions among kinds of goodness, are more obscure: notions that only a philosopher would love. Others, like duties and pain, are familiar to us from everyday life. The normative notion that most of us would find most familiar is one not yet mentioned: the notion of a practical normative reason (hereafter: reason).

It's 7:05 a.m.—time for John to leave for work. He looks outside. It's raining. He takes an umbrella before stepping out.

[2]There is some debate over whether functional uses of "good" are normative—or, as Korsgaard prefers, "normatively loaded" (Korsgaard 1996, p. 42).

Isabella turns on HBO for the next Game of Thrones episode. Her phone starts ringing. It's Kaila. She's saying that she was just laid off. Isabella sets aside her plans for the evening to listen.

Priam looks out past the ramparts of Troy. The Greeks, it seems, are stirring and about to advance. He starts shouting orders to his generals.

"It's raining outside", "My friend lost her job", "They're about to invade"—all of these are reasons. This dissertation is about them. It's about what they are, how our thoughts about them feature in our practical deliberation and activity, how they fit into a larger philosophical picture of our world, and whether it makes sense for us to care about them.

There's a sense in which this kind of dissertation topic is curious. Dissertations are supposed to advance the current state of knowledge of some academically respectable topic. Reasons, however, are already deeply familiar to us. How does one go about expanding the current state of knowledge—at dissertation length—about something that is already deeply familiar? What more is there to know?

Like most philosophers, my answer is that we can expand our knowledge of deeply familiar things, such as our beliefs about reasons, by *articulating* them—or, to use Robert Brandom's colorful phrase, by making these beliefs explicit (Brandom 1998). The articulation of the familiar has always been a particular task of philosophers, and it's a valuable one, both personally and theoretically. Personally, articulating our pre-reflective beliefs is part of fulfilling the Delphic imperative to know ourselves—an important task since, as Xenophon remarks,

> Through self-knowledge men come to much good, and through self-deception to much harm. For those who know themselves, know what things are expedi-

> ent for themselves and discern their own powers and limitations. And by doing what they understand, they get what they want and prosper... (Xenophon 2015, Mem. 4.2.26)

Theoretically, the project of articulating deeply familiar ideas and beliefs is an essential component of the task of determining whether our ideas and beliefs are rationally defensible, and therefore possibly true. While our philosophically pre-reflective beliefs have many virtues—we can use these beliefs to get around in the world, to make predictions, to talk to each other in everyday contexts, and so forth—they also have many vices. Our pre-reflective beliefs tend to exemplify a logical problem: they're frequently inconsistent with each other, which means that some of them must be false.[3] Take my students as an example. If they are moral nihilists, as they frequently say they are (even after I explain to them what the view is), they think that no actions are morally wrong. If they think that engaging in domestic violence is morally wrong,[4] as they always say they do, they must think that there are some actions that are morally wrong. The conjunction of both claims generates a logical contradiction. Something in our pre-reflective beliefs has to go.

Second—and for our purposes, most importantly—our philosophically pre-reflective beliefs can be false for substantive philosophical reasons. Our thoughts about reasons, again, provide what many philosophers would take to be a plausible case in point. Reasons do seem philosophically unusual. They're not ordinary concrete objects. If they're among the furniture of the world, they're a queer furniture indeed. Reasons are not a substance or a material. Objects are not *composed* of reasons. You can't touch or taste a reason—

[3]See (Sverdlik 1985) for a particularly good discussion of inconsistencies between everyday moral beliefs.

[4]And that some actions are in fact instances of domestic violence.

or at least, you can't touch or taste the reason's *being a reason* as, say, we can see an object's being colored.[5] Unlike numbers, reasons don't appear in the sorts of models that one constructs in empirical disciplines like economics, physics, and climate science. So if, as many of us say, the best way to know the world is to defer to our best scientific theories, numbers, philosophically opaque though they are, seem to stand on firmer footing. In the class of things that philosophers take a special interest in, reasons present problems of their own.

1.2 Introducing metaethics

Our prereflective beliefs about reasons fall within the domain of practical philosophy, which we also call ethics. Joseph Raz notes that

> Practical philosophy includes both a substantive or "evaluative" part and a formal part concerned with conceptual analysis. Substantive practical philosophy includes all the arguments designed to show which values we should pursue, what reasons for action should guide our behavior, which norms are binding, etc. The conceptual analysis concerns the logical features of concepts like value, reason for action or norm and the nature of rules of inference governing practical reasoning. (Raz 1975, p. 10).

The distinction Raz is drawing in this passage is a foundational distinction in ethics. When we are asking questions like "should we pull the switch in a trolley case?" (Foot 1967) or "should I engage in commercial surrogacy?" (Anderson 1990), we are asking substantive ethical questions. In answering these questions, we're likely to list some of the reasons for pulling the switch (or not) and engaging in commercial surrogacy (or not): "a train

[5] Even the moral perception literature seems to lean in this direction with its emphasis on the perception of value properties rather than the property of being a reason. See, e.g., (Cullison 2010).

is hurtling toward five hapless victims", "there's someone on the other track", and so on. Substantive ethical questions therefore raise questions about what *the* reasons are. This "what are the reasons?" question is a question about how to construct a special list, one that includes all the reasons and leaves out all the non-reasons. More specifically, they ask the question that this chicken is asking in a motivational poster that the school administration has decided to display at a local middle school:

This chicken's question is not the primary topic of this dissertation. The dissertation concerns, rather, questions in the "conceptual analysis" part of practical philosophy that Raz identifies. While questions about what *the* reasons are falls under substantive practical philosophy, my dissertation concerns the conceptual question of what *reasons* are. When we ask what reasons are, we are not asking for an inventory of the reasons that are out there. We are asking questions about their nature or essence, about what makes something a reason rather than something else.

This general type of distinction exemplified by the distinction between what the reasons are and what reasons are is common in philosophy. Take metaphysics, for instance. One branch of metaphysics, which we call ontology, concerns itself with Quine's question "what is there" (Quine 1948) and wonders such things as, "are there tables?" and "are there

uncountable infinities?" Another branch of metaphysics, whose subject matter Aristotle isolates in Metaphysics E, studies what it is to be, that is, the nature of being itself, or being as being.[6] Similarly, we can distinguish between questions like what counts as money and what is it to be money (Engel 2015), what the virtues are and what it is to be a virtue, and so on.

More closely related to my own concerns in this dissertation is another token distinction of this type that Moore drew over a hundred years ago in his *Principia Ethica* (Moore 1993). Moore had written the first part of the book out of a concern that ethicists, whose business it is to "give reasons for thinking that our statements about... the morality of actions are true or false" (ibid., p. 53), had hitherto ignored another kind of question that had equal claim to falling under the domain of ethics: "what is it that is thus common and peculiar" (ibid., p. 53) to these judgments about the truth and falsity of moral claims? That which is common to these claims, Moore thought, was that they were claims about objects and actions as having a certain property: the property of being *good*. When we ask *what it is* that is common and peculiar to our moral judgments, we are therefore asking questions into the nature of goodness. Other ethicists, Moore thought, had been doing ethics without sufficient awareness of what it is that they were doing.

Moore's particular conclusion about the nature of the goodness property was that it was both *sui generis* and *simple*.

The goodness property was "sui generis", for Moore, in that it was neither a "natural" property nor a "metaphysical" property. Natural properties, for Moore, were those, roughly, that we are used to observing around us—yellowness, roundness and so forth—

[6] (Varzi 2011) draws a similar distinction.

while “metaphysical” properties, in Moore’s idiosyncratic lexicon, were properties of spiritual things, things that we are *not* used to observing around us. Goodness, in being neither a natural nor a “metaphysical” property, occupied a category all of its own, hence its sui generis nature.

By its being “simple”, Moore meant that goodness was not a complex property. Some properties, like my property of being between 5’11” and 6’0”, are complex: they’re instantiated just when their conjuncts are.[7] I instantiate the property of being between 5’11” and 6’0” just if I instantiate the property of being taller than 5’11”, *and* I instantiate the property of being shorter than 6’0”. Goodness, Moore thought, was a simple property, which is to say that conjunction was not among its instantiation conditions.

That goodness was a property *un*like any other kind of property, that goodness was *not* conjunctively complex—these are negatively defined features of goodness. When it comes to positive claims about the nature of goodness, Moore was much more reticent, and, indeed, outright hostile. Moore derides ethicists who attempt to give positive claims about the nature of goodness, accusing them of committing the “naturalistic fallacy”: the mistake of trying to characterize goodness, a sui generis property, in naturalistic terms.

Moore, I remarked, was moved to ask questions about the nature of goodness out of a concern that ethicists, in focusing exclusively on questions about the truth and falsity of claims about what is good and what is not, were not sufficiently aware of what they were doing when they answered these questions. Given the absence of positive claims about the nature of goodness in Moore’s account, one might, I think, reasonably wonder

[7]Properties can be complex without being conjunctive. Disjunctive properties are a case in point. Moore, however, seems to have in mind only conjunctive properties when he discusses complex properties.

whether ethicists, having accepted Moore's purely negative conclusions of the nature of goodness, were in any better position to understand what they were doing when they did ethics. Despite any reservations about the truth or utility of Moore's own conclusions about the nature of goodness, however, the original distinction that motivated his investigation—the distinction between what the moral truths are and what moral truth consists in—is the most important distinction in all of metaethics. The importance is both historical and disciplinary. Disciplinarily, the question of what moral truth consists in is of first importance, as it delineates the terrain that metaethics studies. Historically, as Baldwin notes, "twentieth century British ethical theory is unintelligible without reference to *Principia Ethica*" (Baldwin 1990, p. 66). Nearly all theories of the metaphysics, semantics, and epistemology of normative properties in the early twentieth century can be understood as either extensions of Moore's theory of normativity or as positions that defined themselves against one or more aspects of Moore's theory of normativity.

In the many decades since Moore's contributions to ethics, questions about the nature of goodness and other normative properties have generated not only a library of written work but an entire field of philosophy, one which can be characterized as the field of philosophy that takes as its domain of study the philosophical presuppositions and commitments of our normative thought, talk, and activity.[8] This area has come to be called "metaethics".

An odd name, to be sure, and I think a misleading one, for two reasons. The term "metaethics" conjures up images of a higher-order ethical discipline that takes a lower-order discipline as its subject matter. But metaethics is not an ethics of ethics. It doesn't

[8]This definition is adapted from that given at (Sayre-McCord 2007).

ask, for instance, how one ought to behave at ethics conferences.[9]

Nor, plausibly, is it higher-order in some other sense. Timothy Williamson recently wrote a book in metaphilosophy. He chose to title the book not *Metaphilosophy*, but *The Philosophy of Philosophy*. The reason was, as Williamson says, that "the philosophy of philosophy is automatically part of philosophy just as the philosophy of anything else is, whereas metaphilosophy sounds as though it might try to look down on philosophy from above, or beyond" (Williamson 2008, p. ix). This "outside perspective" view of metaphilosophy is a mistake, Williamson thinks, since metaphilosophy has just as much a claim to being philosophy as all the more standard philosophical specializations. Metaphilosophy is not at a distance from philosophy but rather continuous with it.

Metaethics, too, is more continuous with ethics than the term "metaethics" connotes. Many metaethical questions seem to get their grip not after one steps back to occupy a vantage point from which one can overlook the whole of ethics, but from within the first-order study of ethics itself. Just one example of this is Aristotle who, after asking in Book II of the *Nicomachean Ethics* whether a virtuous character is something we can take steps to cultivate, moves on in Book III to explore the nature of voluntary action. And just as ethical questions can engender metaethical ones, metaethical study can exercise (or at least ought to exercise) a substantial impact on one's substantive normative ethical views. If it's true, as Joshua Green argues, that deontological judgments flow from psychological events such as feelings of disgust that we would rationally judge not to reliably track ethical truth (Greene 2007), then this ought to weaken our commitments to deontological ethical views.

[9]This would be a laudable candidate question for philosophical study, but it's one that tends to be discussed at the *Leiter Report* and *Daily Nous* moreso than in *Ethics* and *Philosophy and Phenomenological Research*, where the magic of metaethics tends to happen.

I'm not sure when the term "metaethics" began to exercise traction over the discipline, although it does start to appear in print in the early 1950s in (Mothersill 1952) and (Wick 1953). It seems to me that a term like "philosophy of practical normativity", when compared to that of "metaethics", would better serve the role of describing the domain of inquiry that it denotes. But "metaethics" does have two virtues that "philosophy of practical normativity" lacks. First, it is familiar to readers; secondly, it is catchy and short. So in this dissertation, I'll set my concerns aside and adopt a terminology that is more amenable to the discipline's discursive conventions.

1.3 Introducing reasons fundamentalism

Systematizing normative reality

Some philosophers find great delight in the huge proliferation of normative thoughts and concepts. One example is Bernard Williams, who eloquently and quite reasonably remarks, in the introduction to his *Ethics and the Limits of Philosophy*, that "we need as many concepts to describe the ethical life as we find we need, and no fewer" (Williams 1985, p. 17). Other philosophers, not sharing Williams' moderate philosophical anarchism, prefer to attempt to provide a systematic theory of the panoply of normative concepts. A common strategy is to isolate a small number of particular normative concepts and then offer purported characterizations of large portions of the other normative concepts in terms of these select few. This is the strategy of, for example, Allan Gibbard's norm-expressivism, which takes norm-acceptance as the central normative concept; Mark Schroeder's Humean theory,

which takes the notion of a desire as central; and T.M. Scanlon's reasons fundamentalism, which places the notion of a reason at the theoretical center. This dissertation focuses on the reasons fundamentalist approach to systematizing normative reality.

Although Scanlon coined the term "reasons fundamentalism" in his (Scanlon 2014), and Scanlon's view is the one that I'll be appealing to the most in this exposition, the reasons fundamentalist view had been elucidated and defended in print by Joseph Raz almost 30 years before. Raz's contention, quite controversial at the time, was that, of all the concepts implicated by our understanding of action and normativity, the concept of "a reason for action is the most fundamental" (Raz 1975, pp. 11-12).

Perhaps unsurprisingly given his background in legal theory, Raz was most concerned in his *Practical Reason and Norms* to give a theory of the normativity of *rules* in terms of reasons for action. As he states in his introduction, "The key concept for the explanation of norms is that of reasons for action. To my mind the main difficulty in explaining rules [rules and norms are used interchangeably] is to understand their relations to reasons for action" (ibid., p. 9). However, by 1999, Raz had adopted a stronger thesis: that "The normativity of all that is normative consists in the way that it is, or provides, or is otherwise related to reasons" (Raz 1999, p. 66). This bold, pithy statement encapsulates a thesis that is at the heart of the contemporary reasons fundamentalist research program.[10]

[10] And, more generally, at the heart of what Wedgwood calls the "reasons-first research program" (Wedgwood 2015). There is currently a similar attempt at systematization playing out in metaphysics. One of the central concepts in contemporary metaphysicsis the concept of grounding. A set of propositions $Q_1, ..., Q_n$ grounds *P* just if its being the case that *P* consists in nothing more than its being the case that $Q_1, ..., Q_n$ (Fine 2001, p. 15). Grounding is of particular interest in part because there is reason to believe that the notion of grounding is fundamental to our understanding of the nature of reality. Some philosophers harbor a hope that this concept will eventually be shown to elucidate other central concepts in philosophy: realism, truthmaking, essence, metaphysical possibility, and so on (Fine 2012, pp. 40-46) (Fine 1994) (Fine 2001). If contemporary metaphysicians manage to pull this off, they will have done a lot to bring these concepts together into something that better resembles a unified conception of the nature and structure of reality. That

Varieties of fundamentality

By stating that normativity of all that is normative consists in the relationship of the normative to reasons, Raz suggests that reasons are fundamental in two ways. First, reasons are the fundamental *source* of the normativity of that which is normative. Consider, for example, these two rules of etiquette: that you shall wait to start eating until everyone has been served, and that peas shall be eaten with a fork. The first (or so I tend to think) has normative force, the second typically does not. The difference, according to reasons fundamentalism, is due to the presence or absence of reasons to follow these rules. Waiting until everyone else is served indicates that we are concerned about the well-being of the other people at the table—something we have reason to communicate to those around us. The rule about eating peas with a fork, in contrast, is an archaic rule of etiquette that most people haven't ever heard of, and is followed so infrequently that we no longer have any reason to do so in any but the most esoteric contexts.[11] Call the *normative fundamentality* of reasons the view that reasons are the fundamental source of normativity.

If reasons are the fundamental source of the normativity of that which is normative, correct *explanations* of the normativity of that which is normative will appeal to reasons. The fact that waiting to eat shows concern for the well-being of the other people at the table not only cites the particular reason that provides the normativity of the rule that you wait to eat. It also explains why the rule to wait to eat is normative. For the reasons fundamentalist, however, the normative nature of reasons neither "needs, or can be given, a philosophical

is a goal philosophically well worth pursuing.

[11]Derek Parfit is a notable exception, having not only heard of the rule but mentions it at (Parfit 2011, p. 145). This passage from Parfit also features a discussion of the distinction between rules with normative force and rules without normative force that I am discussing here.

explanation in terms of some other, more basic notion" (Scanlon 1998, p. 3). Call this the *explanatory fundamentality* of reasons.

In his recent contributions to the reasons fundamentalist view, Scanlon has identified a third way in which reasons can be fundamental. Scanlon has influentially argued that value properties such as goodness are merely formal, second-order properties: "Goodness is not a single substantive property which gives us reason to promote or prefer the things that have it. Rather, to call something good is to claim that it has other properties (different ones in different cases) which provide such reasons" (ibid., p. 11). Value properties, Scanlon argues, are "purely formal, higher-order properties of having some lower-order properties that provide reasons of the relevant kind" (ibid., p. 97). This view is commonly referred to as the *buck-passing account* of value.

On the buck-passing account of value, what it is to have the property of being valuable in a certain way *just is* to have properties that provide reasons of a certain kind. What it is for x to be admirable is for $x$to have properties that provide reasons to admire x; what it is for x to be fearful is to have properties that provide reasons to fear x. Values, on this buck-passing account, are constituted by the property of providing reasons; and reasons are therefore more fundamental than value properties in the sense that they constitute value properties. Call this the *constitutive fundamentality* of reasons.

The constitutive fundamentality claim entails the explanatory fundamentality and normative fundamentality claims. If values, for instance, are just higher order properties of having lower order properties that provide reasons—if this is what it is to be a value—then values could only derive their normativity *through* reasons. And if there is no further fact from which reasons derive their own normativity, values could only derive their normativ-

ity *from* reasons. The mere higher-orderness of a property provides neither an additional source of normativity nor an additional explanation of it; there's nothing else to provide the normativity of values except the reasons.

Reasons, for Scanlon, are constitutively fundamental to a variety of normative properties, not simply the evaluative ones. Scanlon understands wrongness, for example, in terms of reasonable rejection of principles:

> judgments of right and wrong by saying that they are judgments about what would be permitted by principles that could not reasonably be rejected, by people who were moved to find principles for the general regulation of behavior that others, similarly motivated, could not reasonably reject. In particular, an act is wrong if and only if any principle that permitted it would be one that could reasonably be rejected by people with the motivation just described (or, equivalently, if and only if it would be disallowed by any principle that such people could not reasonably reject). (ibid., p. 4)

An historical aside

Raz's *Practical Reason and Norms*, along with Nagel's *The Possibility of Altruism* (Nagel 1978), launched, if not a research program, at least a huge surge in interest among analytic philosophers in the notion of a reason for action. According to John Broome, the 1970s was the "age of the discovery of reasons" (Broome 2004, p. 24). And Broome is not the only one who seems to view the intellectual history in this way. Parfit, for example, goes as far as to claim that reasons were not much discussed before the 70s because the older generation of moral philosophers, and Bernard Williams in particular, "did not understand [the] concept of a reason" (Parfit 2011, p. 434).

Two curiosities lurk in the background here. Irrespective of whether Parfit is right

about what concepts Williams understood,[12] it's certainly true that there was, comparatively speaking, little discussion of reasons before the 70s. Suppose for a second that Raz is right about the conceptual fundamentality of reasons. How could moral philosophers have been so completely in the dark, for so long, about the very thing that Raz is claiming is fundamental to the most important concepts that constitute our understanding of practical reality?

I have no response to this question. It is a real mystery. It is like asking why Aristotle never dropped two rocks of different masses off the Parthenon and saw that his theory of gravity is false. Why, Aristotle? How could you not do that?[13]

For the second curiosity, what about Anscombe's influential work on reasons in her magisterial *Intention*, which was published in 1957? Why don't Broome and Parfit fail to locate the age of the discovery of reasons a full two decades earlier? You can't chalk it off to the book's general lack of influence, since *Intention* frequently makes it on top ten lists of most influential philosophical books of the 20th century (Ho 2009), (Leiter 2010).

I have not one but two responses to this curiosity: one philosophical and one sociological.

Philosophically, the kind of reasons that play the star role in Anscombe's *Intention* are *motivating reasons*: reasons why one *did* something. Motivating reasons are not (or at

[12] I remember Parfit saying in seminar that he was driven to this view out of respect for Williams' philosophical acumen. Parfit, as I remember it, said that he had difficulty accepting that a philosopher as adept as Williams could have been led to accept the internalist conclusions about normativity that Williams argued for in "Internal and External Reasons" (Williams 1981, pp. 101-113), and that Parfit thinks are decisively false.

[13] Even more mysterious: why did no one do this for 2000 years until Galileo showed up? One response: they did, but they didn't write about it because of the influence of Aristotle on the Church and the Church's power over anyone who could write. Analogously, perhaps plenty of philosophers had the concept of a normative reason for action before the 1970s, but they didn't write about it because of the influence of overly scientistic conceptions of the world on the most influential philosophers at the time and these philosophers' power over anyone who wanted to publish in philosophy journals. This is just idle speculation.

least need not be in every case) *normative reasons*, which are reasons why one *ought to have done* something. *Intention* does not discuss normative reasons. They had yet to be discovered.

Sociologically, Anscombe's work on intentional action has been hugely influential, but the work has exercised a greater influence over more recent philosophers of action. David Velleman, Michael Bratman, Carl Ginet, and Michael Thompson, all of whom have contributed substantially to our philosophical understanding of action, often and explicitly acknowledge Anscombe's influence on their thought and work. In contrast, the most prominent metaethicists do not explicitly recognize their debt to Anscombe in the same way. Scanlon cites Anscombe less than a handful of times in his authoritative (Scanlon 1998), and in every case the reference is to an aspect of Anscombe's work that relates only indirectly to the topic of reasons for action. Similarly, Parfit limits his citation and discussion of Anscombe to the parts of his *On What Matters* that deal with normative ethics. Scanlon and Parfit both seem to be writing within an intellectual tradition in which Anscombe's influence on action theory and the theory of reasons is not as recognized. Parfit might have simply forgotten about Anscombe's work on reasons when he writes that, before Raz and Nagel, no one in philosophy had been talking about reasons.[14]

[14]Parfit is charging Williams with *not grasping the concept of a normative reason*, so I think it's not uncouth for me to charge Parfit with the much less egregious mistake of failing to recall Anscombe's work on reasons. Furthermore, Parfit is famously absent-minded (Macfarquhar 2011), so the charge is not implausible. Broome falls distinctly within the Scanlon-Parfit tradition in virtue of his heavy interactions with Parfit, so it is not surprising from a sociological point of view at least why Broome's only mention of Anscombe in (Broome 2004) is relegated to a footnote on page 24 in which he explicitly thanks someone else for reminding him about Anscombe.

1.4 The thesis

This dissertation argues for a three-part thesis. First, this dissertation argues for a particular interpretation of reasons fundamentalism. According to reasons fundamentalism, the property of being a reason is a relational, irreducibly normative, and primitive property whose distribution supervenes on the distribution of non-normative properties. By irreducibly normative, I mean not identical to any purely descriptive property; and by primitive, I mean not decomposable into simpler properties. The first part of my thesis, which I offer in chapter 2, argues for a particular precisification of this general characterization.

Second, this dissertation argues that although reasons fundamentalism, on my interpretation, holds that the normative features of the world supervene on the world's descriptive features, reasons fundamentalism cannot explain why the normative and descriptive features should relate in this way. This argument plays out across chapters 3 and 4.

Third, this dissertation argues for a novel metaethical theory that preserves key virtues of the reasons fundamentalist view while also answering the explanatory challenge. On this theory, the normative notion that best systematizes the panoply of metaethical concepts is not the notion of a reason but the notion of a genuinely thick property, that is, a property that has descriptive and normative content such that these contents cannot be broken apart into components that are all purely normative or purely descriptive. On my theory, genuinely thick properties are similar to the reasons fundamentalist construal of reasons in three ways: they are irreducibly normative; they are primitive; and they partially constitute thin properties. Genuinely thick properties, I argue, do not supervene on non-normative properties, and in this way they are unlike the reasons fundamentalist conception of the relation

between reasons and the descriptive features of the world. The genuinely thick properties, however, do constitute a supervenience base for other key normative properties such as thin properties and the property of being a reason; and these supervenience relations can do all the explanatory work that metaethical theory needs supervenience claims to do. I present this view and the related arguments in chapter 5.

Chapter 2

Reasons Fundamentalism

This chapter is a detailed exposition and articulation of reasons fundamentalism. Most broadly, reasons fundamentalism is the view that there are irreducibly normative truths about reasons for action (Scanlon 2014, p. 1).More specifically, reasons fundamentalism is particular conjunction of views defended by Raz, Parfit, and Scanlon. These conjuncts we can call the *Centrality*, *Relationality*, *Primitivity*, *Irreducibility* and *Additive Reasons Supervenience*conjuncts. I'll go ahead and define these conjuncts here, so that my readers have a handy place to look them up. However, they're going to need a lot of explaining. So, after listing the conjuncts of reasons fundamentalism here, I'll unpack the conjuncts in more detail in the rest of the chapter.

The conjuncts:

Reasons Centrality: For all facts F, F is a normative fact only if F is a reasons fact or F embeds a reasons fact.

Reasons Relationality: Any reason fact is normative in virtue of its standing in the favoring relation with a circumstance-agent-action triple.

Reasons Primitivity: The favoring concept is primitive.

Reasons Irreducibility: The favoring relation is not normatively masked.

Additive Reasons Supervenience: For any possible worlds w_1 and w_2 and any fact-

agent-action-circumstance tuples T_1 in w_1 and T_2 in w_2, if T_1 and T_2 are intrinsically and extrinsically non-normatively indiscernible and w_1 and w_2 are non-normatively indiscernible, then T_1 instantiates the favoring relation if and only if T_2 instantiates the favoring relation.

Typically, reasons fundamentalism is also construed as also offering a theory of rational motivation, of how reasons motivate actions. Despite this, I will not consider rational motivation in this chapter. The reason for this is because rational motivation is a question that appropriately belongs to moral psychology and is therefore at least somewhat orthogonal to my own area of investigation: the metaphysics of reasons. I will, however, be in a position to discuss rational motivation at the very end of this dissertation, after I have discussed in considerably more detail the role and nature of thick normative properties.

I should mention that the exposition of reasons fundamentalism that I give in this chapter, both in its general presentational features and in the substantive details, is my own. As far as I know, no one has described the theory as the conjunction of the above six claims; and no one, as far as I know, has formulated the conjuncts in the way that I have. In particular, my presentation of reasons fundamentalism is not Scanlon's presentation. However, I take it that the substantive theory I describe in this chapter is a proper part of Scanlon's theory in all key respects, and differs in presentation only because I felt that it was important to give a more precise characterization of various features of Scanlon's view in order to adequately formulate and develop the arguments of §3 - §5 of my dissertation. Throughout the presentation, I also mention ways that my formulation of reasons fundamentalism captures certain key features of Parfit Raz, and Enoch's theories as well.

Furthermore, the philosophers who assert the six conjuncts I have just mentioned also hold a variety of other positions on the nature of normative reality. I am specifically using "reasons fundamentalism" in this chapter to refer to the conjunction of the six philosophical theses I have just enumerated. For example, Scanlon argues that there exist *pure normative truths* such as "pain is bad" that hold with metaphysical necessity and provide principles of explanation for why any particular reason is a reason. In §4, I argue against the idea that there are pure normative truths. However, although my argument against the existence of pure normative truths in that section is an argument against one of Scanlon's positions, it is not an argument against reasons fundamentalism on my construal of reasons fundamentalism, since the existence or nonexistence of pure normative truths is an issue that turns out to be independent of the truth of the six philosophical theses I have just enumerated. I mention this only to clarify that the reasons fundamentalist position that I describe in this chapter is specifically the conjunction of Centrality, Relationality, Primitivity, Irreducibility, Additive Reasons Supervenience, and Non-Humean Normative Motivation. That's it. Although I do at times discuss further aspects of Scanlon's theory, my exposition in this chapter is not intended to be an exposition of Scanlon's entire theory of normativity.

2.1 Centrality

Centrality explained

Here again is the Centrality conjunct of reasons fundamentalism:

> *Centrality*: For all facts F, F is a normative fact only if F is a reasons fact or F embeds a reasons fact.

To explain this, I'll first discuss how I'm using the term "fact". Then I'll define "reasons fact". Lastly, I'll explain what it means to embed a reasons fact.

I understand facts to be entities in which particulars exemplify properties or stand in relations.[1] In this I agree with Baylis when he writes,

> Facts are entities in relation. At the lowest level they are relations among particulars such as one patch or one building being larger than another, or they are relations of exemplification between particulars and the characters which characterize them, as for example a given patch being red. (Baylis 1948, p. 462)

Facts are recursive: if it is a fact that Obama likes Predator drones, then it is a fact that it is a fact that Obama likes Predator drones, it is a fact that it is a fact that it is a fact that Obama likes Predator drones, and so on. The recursion of facts evidences that there can be facts about facts: for instance, the fact that it's a fact that Tim Scanlon is a philosopher, or the fact that the fact that we have reasons not to commit terrorist acts is a normative fact. We can define a *factual embedding* function recursively as follows:

> For any facts F and G, if F is a fact about G, then F *embeds* G.
>
> For any facts F, G, and H, if F embeds G and G embeds H, then F embeds H.

Reasons, for Scanlon and most people who work on reasons, are facts. A *reasons fact*, as I'm using it, is just a fact that some other fact is a reason. So when Centrality says that any normative fact is either a reasons fact or a fact that embeds a reasons fact, it is saying that any normative fact is either a fact that something is a reason, or a fact about the fact

[1] I slightly modify this formulation from (Correia and Mulligan 2013). At times, I may talk of sets of particulars "exemplifying" some relation or relational property R. If I say this, I mean to say that these particulars stand in R. At times, I may also talk of some x as being an "object" exemplifying a property or standing in some relation. If I say this, I mean to say more broadly that x is a particular that exemplifies the property or stands in the relation.

that something is a reason, or a fact about a fact about a fact that something is a reason, and so on.

One potentially surprising implication of Centrality is that most reasons are not normative facts, since most reasons are not facts about some other facts' being a reason.[2] Take, for example, the fact that air pollution causes global warming. On the reasons fundamentalist way of thinking about things, the fact of air pollution's causing global warming is not in itself a normative fact: there's nothing about *what it is* for air pollution to cause global warming that entails anything about what we should do or how the world should be. For the reasons fundamentalist, it's only *the fact that* the fact that air pollution causes global warming *is a reason* that in itself entails anything about what we should do or how the world should be.

When Centrality states that for all facts F, F is a normative fact only if F is a reasons fact *or F embeds a reasons fact*, it is allowing for there to be normative facts other than facts about certain facts being reasons. However, it requires that any such normative facts be about facts about certain facts being reasons.[3]

To illustrate this aspect of the Centrality principle, we can take another look at Scanlon's buck-passing account of value. On the buck-passing account of value, for a particular individual x to be valuable is for x to have a feature P such that x's having (exemplifying) P is a reason to value (promote) x. The fact of P's being valuable consists in nothing over and above the fact that the fact that P's having F is a reason exemplifies a certain property:

[2]Some reasons are. The fact that terrorism kills innocent people is a reason not to commit acts of terrorism. But the fact that terrorism's killing innocent people is a reason *is itself a reason* for me to use it in this example.

[3]This is probably confusing, but I don't think that that's my fault. It has to do with the fact that the Centrality conjunct of reasons fundamentalism makes claims not only about normative facts but also about facts about normative facts and facts about facts about normative facts.

the property of being a reason *to value* P.

Here is an example. Suppose (counterfactually or not) that philosophy is fun:

Fact 1: FUN(philosophy).

Suppose further that *Fact 1* is a reason:

Fact 2: REASON(FUN(philosophy)).

Fact 2 can exemplify lots of properties, such as the property of being a reason for loving philosophy, or for admiring philosophy, and so on. Suppose that the funness of philosophy is a reason to *value* philosophy—that is to say, that the fact that the funness of philosophy is a reason has the property of being a reason fact of the valuing kind:

Fact 3: VALUING_KIND(REASON(FUN(philosophy))).

On the buck-passing account of value, for philosophy to be valuable just is for philosophy to have a feature that constitutes a reason to value philosophy. In this case, that feature is philosophy's funness. So the buck-passing account of value is this:

Buck-Passing Value Definition:
$\forall x[\exists Q(Q(x) \wedge$ VALUING_KIND(REASON(Q(x))))
$\leftrightarrow$ (VALUABLE(x))].

which, along with *Fact 1* and *Fact 3*, gives us

VALUING_KIND(REASON(FUN(philosophy))) $\rightarrow$
VALUABLE(philosophy).

The fact that philosophy is valuable is a normative fact. But it is not a reason fact, since it is not a fact that anything is a reason. The reason fact in our example is *Fact 2*. However, although the fact that philosophy is valuable is not a reason fact, it does embed a reason fact, as we can see in *Fact 3*. So Centrality allows the fact that philosophy is valuable to be a normative fact as well.

2.2 Relationality

Raz's early work articulated not only the Centrality principle but also a view as to what the reasons are, and what makes them so. As Raz notes,

> We usually think of reasons for action as being reasons for a person to perform an action when certain conditions obtain. The performance of an action by an agent in certain circumstances can be regarded as a fact, and it may be thought that reasons are relations between facts. (Raz 1975, p. 19)

In this short paragraph, Raz is saying two significant things. Reasons, first of all, are facts. But not all facts are reasons, which raises the following question. Take all the reason facts in your left hand, and all the non-reason facts in your right. What's the basis of dividing up the facts in the way that we just did? Raz's answer is very simple, and yet, for all its simplicity, has been exceedingly helpful in advancing our philosophical understanding of reasons. What accounts for reasons' being what they are is that the facts that are reasons possess a certain property—the reasons-making property—that makes them so. This property, furthermore, is a relational property. Facts that are reasons bear the reasons-making relation toward a person *for whom* the fact is a reason, the action *for which* the reason is a reason, and the set of conditions *in which* the fact is a reason for that person to do that ac-

tion. Since the reasons-making property is a relational property, the literature has taken to calling the property the "reasons relation".[4] Call any tuple *"reasons-apt"* just if it's a fact-agent-action-circumstance tuple—in other words, that it's of the right form for instantiating the reasons relation.

It's now widely accepted that the reasons-making property is a relational property. The most outspoken contemporary defender of Raz's view[5] that the reasons relation is a four-place relation is Scanlon, who writes:

> Whether a certain fact is a reason, and what it is a reason for, depends on an agent's circumstances. The fact that this piece of metal is sharp is a reason for me not to press my hand against it, but under different circumstances it might be a reason to press my hand against it, and under still different circumstances a reason to do something else, such as to put it into the picnic basket if I will later have reason to want to cut cheese. This suggests that "is a reason for" is a four-place relation, $R(p, x, c, a)$, holding between a fact p, an agent x, a set of conditions c, and an action or attitude a. This is the relation that holds just in case p is a reason for a person x in situation c to do or hold a. (Scanlon 2014, pp. 30-31)

The reasons-making relation is called the *favoring relation*, a term which again we owe to Scanlon.[6]

The difference between reasons and the reasons-making relation is more important than it may at first seem. Scanlon and Raz identify reasons with facts. So reasons will be philosophically problematic in the same ways that facts are.[7] However, reasons pose philosoph-

[4] Peacocke sometimes refers to the relation as the "reasons for" relation. I don't use this particular locution.

[5] Although, it seems, without citing Raz.

[6] I had actually thought that Dancy had originated the term, but Dancy cites Scanlon as his inspiration at (Dancy 2004, p. 29). Not everyone understands the favoring relation to be four-place. Dancy, for example, talks of it as being a two-place relation. The four-place structure of the favoring relation is a commitment of reasons fundamentalism in particular, not of reasons realists, such as Dancy, in general.

[7] The issue regarding facts that most easily comes to mind is not a terribly recent one: Davidson's Slingshot Argument. I've found (Soames 2008) and (Merricks 2007) helpful in understanding the current state of this debate.

ical issues over and above those posed by facts. The locus of these additional issues will be, not the reasons themselves—for these are just facts on Scanlon and Raz's account—but the property in virtue of which the reason-facts are reason-facts and not just facts of some other kind. The shift from reasons to the reasons relation therefore better isolates the philosophical issues that are in the metaethicist's purview.

The normativity of reasons facts derives from the favoring relation they bear to some circumstance-agent-action triple. Take, for instance, the

> *Tyranny Fact*: Acts of terrorism against citizens of Western countries tend to strengthen tyrannical impulses on the part of Western governments.

For reasons fundamentalists, reasons are facts that stand in the favoring relation to a circumstance-agent-action triple. So when I am saying that Tyranny Fact is a reason for critics of the West not to commit terrorist acts in our particular global situation, I am saying that Tyranny Fact stands in the favoring relation to a triple consisting of our current global situation, critics of the West, and the action of committing terrorist acts. In addition to Tyranny Fact, there is also the numerically distinct fact that that Tyranny Fact, our current global situation, critics of the West, and the action of committing terrorist acts all stand in the favoring relation to each other. Since the favoring relation is a normative relation, this latter fact is, in fact, a normative fact. Crucially, however, according to reasons fundamentalism, *Tyranny Fact is not a normative fact*. Any and all of the normative character that Tyranny Fact possesses is derived entirely from its standing in the favoring relation to a circumstance-agent-action triple.

2.3 Primitivity

Introducing primitivity

If the favoring relation is the source of all value, the philosophically minded might well start to wonder: well, what is this favoring relation anyway? The literature redounds with attempts to elucidate the favoring concept in terms of something more philosophically tractable—for instance, in terms of our concept of a pro-attitude (Schroeder 2007), commitment (Street 2008), norm-acceptance (Gibbard 1990, p. 163), explanations of ought facts (Broome 2013, pp. 54-55), and evidence for ought facts (Kearns and Starr 2008). Scanlon's view, however, is that these elucidations all fail:

> I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favor of it. "Counts in favor how?" one might ask. "By providing a reason for it" seems to be the only answer. So I will presuppose the idea of a reason. (Scanlon 1998, p. 17)

Parfit agrees:

> It is hard to explain the concept of a reason, or what the phrase "a reason" means. Facts give us reasons, we might say, when they count in favour of our having some attitude, or our acting in some way. But "counts in favour of" means roughly "gives a reason for". Like some other fundamental concepts, such as those involved in our thoughts about time, consciousness, and possibility, the concept of a reason is indefinable in the sense that it cannot be helpfully explained merely by using words. (Parfit 2011, p. 31)

There are two things I want to say about this. First, the locution that reasons are considerations that "count in favor", which we see in both quotations, is the source of our contemporary proclivity to call the reasons relation the "favoring relation". So if you see the term

"favoring relation" either here and elsewhere, all it means is the reasons relation. Second, by saying that the concept of a reason is "hard to explain", Parfit isn't simply engaging in the overused philosophical trope of saying that some philosophical concept is "notoriously difficult to define".[8] Parfit is not simply saying that it's difficult to define "reason", but that it's actually impossible to define "reason" in a correct, non-circular way.

These passages indicate that Scanlon and Parfit are *primitivists* about the favoring concept. A concept is primitive just if it does not admit of a correct, non-circular analysis. Although Scanlon and Parfit do hold that "favors" means "is a reason for", they hold that these distinct locutions are simply different ways of expressing identical concepts. In claiming that the favoring/reasons-for concept is primitive, Scanlon and Parfit are claiming that this single concept that both locutions express does not admit of any further elucidation in terms of simpler concepts that may be more philosophically tractable.

Scanlon's claim that a reason is a "consideration that counts in favor" of performing a certain action (Scanlon 1998, p. 17) introduces a complication in Scanlon's thought. Scanlon also thinks reasons are facts. However, facts and considerations are distinct. Considerations, for instance, seem to be mental events, while facts are not: they're part of the furniture of the world. A consideration can be of a fact, and there can be facts about considerations,[9] but facts and considerations are nevertheless distinct kinds of things. Are reasons facts, then, or are they considerations? Although this is an important question, I will set it aside in this dissertation and assume, along with Scanlon, that reasons are facts.[10]

[8]For examples of this trope, see most entries in the Stanford Encyclopedia.

[9]And facts about considerations of facts about considerations of facts, and so on—see (Varzi 2012).

[10]In §5, I argue that reasons are answers to why-questions. Answers are typically understood to be propositions rather than facts. However, it is consistent with my arguments in §5 to say, not that reasons are answers

The normative question objection

Discussion of reasons primitivism predates Scanlon and Parfit's defense of it. Gibbard remarks that "Some writers speak of 'reasons' in a non-Humean way, and indeed try to ground ethical theory on a non-Humean concept of reasons... None of them, so far as I can discover, explains what he is using the term 'reason' to mean" (Gibbard 1990, p. 161). Gibbard suggests here that the non-Humeans' general inability to articulate what they take reasons to be, and what they mean when they discuss reasons, to be a deficit of non-Humean theories of normative reasons such as those of Scanlon and Parfit. One major deficit, Gibbard sees, is that, without any elucidation of the reasons concept, the non-Humean cannot explain what it is about reasons that makes them normatively salient.

To put the objection another way, suppose we come to believe that volunteering at the shelter this Saturday is what we have most reason to do. Even armed with this belief, we still have not exhaustively settled the question of deciding what to do. What is it about *having most reason* to volunteer at the shelter this Saturday that settles the question *what to do* on Saturday? Gibbard worries that, without further elucidation of the concept of a reason, which the reasons primitivist can't provide, there remains both an explanatory gap and a practical gap between an action's property of being such that we have most reason to do it, and the action's property of being "the thing to do" or "what to do". Call this the *Normative Question Objection* to reasons primitivism.[11]

---

to why-questions, but that they are such as to provide answers to why-questions or bear some other close relationship to answers to why-questions. So I think that my remarks in §5 do not help to settle the question whether reasons are facts, propositions, considerations, or something else entirely.

[11] These remarks are an adaptation of Gibbard's comments at (Gibbard 2003, pp. 15-16). His remarks there focus on the "ought" property rather than the property of being a reason, and he doesn't mention reasons primitivity explicitly. A good review of literature on the normative question, especially as it relates to reasons fundamentalism, can be found at (Dreier 2015).

Gibbard's remark here is interesting not only because it substantially predates the rise of reasons primitivism, but also because Gibbard on the one hand, and the non-Humeans Scanlon and Parfit on the other, have conflicting approaches to the question whether the failure to provide any elucidation of the concept of a reason is a theoretical virtue or a theoretical vice. Scanlon, for instance, thinks that reasons primitivism is fully capable of answering the explanatory challenge that Gibbard raises. According to Scanlon, my reasons to volunteer at the shelter on Saturday are normative for an agent like, say, me, because, to the extent that there are reasons for me to volunteer at a shelter on Saturday, these reasons are *reasons for me* to *volunteer at the shelter on Saturday* (Scanlon 2014, p. 10). Scanlon thinks that no "further explanations [of why my reasons are normative for me] need or can be given" (ibid., p. 44). If there *were* a live question as to why my reasons to volunteer on Saturday were normative for me, then I think it's correct to say that a reasons realist would face pressure to give some sort of elucidation of these reasons in terms that made more readily apparent the connection between something's being a reason for me, and something's being normative for me. However, Scanlon thinks the question of why my reasons are normative for me is not a live question: it's sufficiently answered in the reasons' status of being *reasons for me*. The question, in other words, admits of a trivial answer and therefore does not pose any philosophical worry. Scanlon concludes that the Normative Question Objection does not provide any reason to give up reasons primitivism.[12]

In the next section, I will investigate Gibbard's concern more thoroughly. However, the

[12]Furthermore, even if the normative question were a serious objection to reasons primitivism, Scanlon charges Gibbard's variety of expressivism with an implicit commitment to reasons primitivism at (Scanlon 1998, p. 58). So reasons primitivism's putative failure to answer the normative question would therefore not be a reason to defect to Gibbard's norm-acceptance theory of reasons.

arguments of this next section not serve any further role in the overall argumentation of this dissertation. Readers pressed for time can therefore easily skip it.

Caring about normativity

Here is another way to state the Normative Question Objection. A number of objections against reasons fundamentalism take this form:

1. The favoring relation possesses feature F.
2. If reasons fundamentalism is true, then the favoring relation does not possess F.
3. Therefore, if the favoring relation possesses F, then reasons fundamentalism is not true. (2, contraposition)
4. Therefore, reasons fundamentalism is not true. (1, 3, modus ponens)

Here's an instance of an argument that takes this form:

1. We *care about* whether things exemplify the favoring relation. (premise)
2. If reasons fundamentalism is true, then we can't say anything informative about what the favoring relation is like. (Primitivity premise)
3. If we can't say anything informative about what some property or relation is like, then we do not care whether things exemplify that property or relation. (premise)[13]
4. Therefore, if reasons fundamentalism is true, then we do not care whether things exemplify the favoring relation. (2, 3, transitivity, universal instantiation)

[13] In other words, we don't care whether things exemplify relations that we can't say anything informative about. This formulation is clearer than the one I provide in the text of the argument, but the formulation that I provide in the text of the argument makes, I think, the argument's validity more apparent.

5. If we care whether things exemplify the favoring relation, then reasons fundamentalism is not true. (4, contraposition.)

6. Therefore, reasons fundamentalism is not true. (1, 4, modus ponens)

Call this the *Descriptive Caring Objection*.[14]

Korsgaard provides a version of the Descriptive Caring Objection when she writes:

> Why after all should we govern our lives, or anyway restrict our private interests, for the sake of the promotion or maximization of things possessing a non-natural property about the character of which we are able to say very little? Is it really enough to say that after all this is the *good* we're talking about, not just any old property? If it is a natural state or object... we might feel somewhat easier about the intelligibility or commitment to it. (Korsgaard 1981, p. 10).

Korsgaard's target here is a primitivist about goodness, but her concerns about primitivist accounts of goodness are easily adapted to target primitivist accounts of the favoring relation. Intuitively, we care about whether some fact or other is a reason—which is to say, whether that fact stands in the favoring relation. But, according to reasons fundamentalism, we are only able to say very little about this relation. Its character is mysterious, and calling this relation the "favoring relation" doesn't help dispel the mystery. Usually, we care about things that are not totally mysterious, like our children and our careers. We can say something about why these things we care about are careworthy. But if reasons primitivism is true, we can't do this about the favoring relation, so it's hard to see how we can care about it. In fact, the reasons fundamentalists' commitment to primitivity means that they can't say anything about how we can care about it. So reasons fundamentalism is false.

I argue that we should reject premise (1) of the Descriptive Caring Objection.

[14] I call it the *Descriptive* Caring Objection, rather than just the "Caring Objection", in order to preemptively distinguish it from a normative version of this objection that I raise in a few paragraphs.

According to Korsgaard, we do not care about governing our lives "for the sake of the promotion or maximization of things possessing an [insurmountably mysterious] property". This is ambiguous. In particular, the sentence admits of *de dicto* and *de re* interpretations. On the de dicto interpretation, the sentence says that we do not care about promoting *what-ever* states of affairs and actions that happen to instantiate the favoring relation, whatever they happen to be. On the de re interpretation, the sentence says that we do not care about promoting *the very* states of affairs and actions that instantiate the favoring relation. In other words, on the de re interpretation, Korsgaard is claiming that we do not care about things that happen to instantiate insurmountably mysterious properties.

Korsgaard's claim is plausible only on the de dicto interpretation. When we are moved by the Paris killings, we are moved by these acts' natural properties: their properties of causing death and destruction to many innocent people. Although it's true that these acts are also wrong, we're not primarily moved by these acts' wrongness. We care far more about these natural properties of the Paris killings. It is not plausible that we do not care about the natural properties of actions like the Paris killings that also happen to instantiate normative properties, which is what the de re interpretation claims.

Tim Chappell argues, I think correctly, that the *de re* reading is more in line with what reasons fundamentalists actually think. If this is true, then the de dicto interpretation of Korsgaard's claim, which I argued is the only plausible interpretation, will not constitute an objection to reasons fundamentalism.

It's surely true that what we care about in practical contexts is the natural world. Instead of claiming that we also care in practical contexts about normative properties, the primitivist can claim that the normative properties serve rather as *markers* of things in the world that

it's correct to care about (Chappell 2014). We care about our children's health, for example; and when we tend to a child who is ill, we are responding to descriptive facts like the fact that my child is ill or that my child is in discomfort. What we really care about is that my child is ill and in discomfort rather than in a healthier state. We don't first and foremost care about whether our child's health has the further property of being valuable. In this, the reasons fundamentalist agrees with Korsgaard.

The reasons fundamentalist will say that there is still the normative fact that alleviating our child's illness or discomfort is correct or valuable or right. This normative fact functions as a truth-maker for our claims that our actions are correct, or valuable, or right. But it can function as a truthmaker without our having to care about it. So reasons fundamentalists can maintain that the favoring relation serves as constituents of truthmakers for normative claims while avoiding commitment to the claim that we care about these truthmakers, thereby avoiding (1) and therefore avoiding (6).

Surely some of us do care whether the actions we perform and states of affairs we promote are normatively favored or are valuable. Parfit, for example, seems to care a whole lot, particularly when he concludes that, if there are no irreducibly normative truths, then he "would have wasted much of [his] life" (Parfit 2011, p. 304).[15] So it is open to the reasons fundamentalist to reject (3) and therefore (4) and (6).

I think that this response on behalf of reasons fundamentalism is a mistake. The mistake turns on an ambiguity in the notion of caring. There is a theoretical kind of caring that we

[15]Korsgaard's response to this would likely be that Parfit is just idiosyncratic, as she says in (Korsgaard 2009). I believe Parfit's response to the idiosyncrasy claim would be that, yes, he is idiosyncratic, but his idiosyncracy is a virtue when nearly everyone else in the field of normative theory has been so horribly wrong about the nature of normativity for so many decades.

have when we are gripped by theoretical questions. But I don't think Korsgaard disagrees that we theoretically care about irreducibly normative truths. We are philosophers and all know, or at least remember, what it's like to be gripped by theoretical questions. In talking about caring, I think she has in mind the practical kind of caring that we are all familiar with. I suspect that Parfit is concerned that he would have wasted much of his life if there were no irreducibly normative truths in significant part because much of his life was spent studying irreducibly normative truths. If there aren't any, then many of the non-irreducibly-normative aspects of his life that he does care about—his career, thousands of his evenings, and so on—would have been spent chasing a theory about something that doesn't exist. I suspect most defenders of irreducibly normative truths—and I am one of them—care about exemplifications of the favoring relation because the value of many of their personal projects and pursuits, which they do care about, would lose much of their value if there turn out not to be any irreducibly normative truths. This is no different than the kind of care that some physicists bear towards theories of quantum gravity, or that computer scientists have for the PNP problem: if there is no theory of quantum gravity or no solution to the PNP problem, their personal projects would lack certain natural features—going down in the textbooks, keeping their jobs, being respected by their academic communities, defeating their theoretical opponents—that they do care about. In the case of moral philosophers no less than physicists or computer scientists, our practical caring about the truth of the theories we create is derived from the care we have for the various practical implications of the success or failure of these theories.

There is another sense in which we care about exemplifications of the favoring relation. When we are deciding what to do, we will ask questions like "is such-and-such a course of

action wrong" or "do the reasons weigh in favor of doing professional philosophy or going to law school". When we are asking these questions, we certainly care about whether and in what ways the favoring relation is exemplified. But I suspect that the care we have for these is again derivative of the non-normative features of our lives and world for which we have practical concern. Reasoning about whether something is wrong or whether something is what we have most reason to do is reasoning about whether certain courses of action have normative properties. But this reasoning often functions as a surrogate that takes the place of reasoning about our cares and concerns directly. If I am of a certain kind of Kantian bent, I might reason in the following way: ϕ-ing is not universalizable, therefore ϕ-ing is wrong, therefore I will not ϕ. But the universalizability method of determining the permissibility or impermissibility of certain courses of action can function as a heuristic for determining whether ϕ has certain other non-normative features that I care about, such as the feature of showing respect for the humanity of the other persons that I'm interacting with, or the feature of being freely chosen, or the feature of being an action at all.[16] The wrongness of ϕ-ing is indeed something that I would care about if I were a Kantian of this bent, but only because of certain non-normative features of ϕ-ing that the wrongness of ϕ-ing entails.

We can put the caring objection in another way, not in terms of a descriptive constraint on theories of reasons, but in terms of a normative constraint:

1. We *ought to care* whether things exemplify the favoring relation. (premise)
2. If reasons fundamentalism is true, then we can't say anything informative about the

[16] I reiterate that I do not think that freedom and actions are non-normative. I am stating that they are non-normative here because it better reflects the reasons fundamentalist position that I am developing. I have more to say on the normative nature of actions in §4.

favoring relation. (Primitivity premise)

3. If we can't say anything informative about a property or relation, then we ought not to care whether anything exemplifies that property or relation. (premise)
4. Therefore, if reasons fundamentalism is true, then we ought not to care whether anything exemplifies the favoring relation. (2, 3, transitivity)
5. Therefore, if we ought to care whether anything exemplifies the favoring relation, reasons fundamentalism is not true. (4, contraposition)
6. Therefore, reasons fundamentalism is not true. (1, 5, modus ponens)

Call this the *Normative Caring Objection* to reasons fundamentalism.

I think a reasonable response to this argument is, again, to deny (1). Nietzsche famously worries that

> Our social conceptions of good and evil, weak and effeminate as they are, and their enormous influence over both body and soul, have had the effect of weakening all bodies and souls and of crushing all unprejudiced, independent, and self-reliant men, the real pillars of a strong civilization: wherever we still find the evil morality today, we see the last crumbling ruins of these pillars. (Nietzsche 2012, §163)

Nietzsche's worry, expressed here, raises a concern that caring about whether normative properties are exemplified may have deleterious consequences for the flourishing of other non-normative things that we happen to care about, such as personal strength, "masculinity", and strong civilizations. If we happen to care more about these non-normative things than about whether our actions exemplify normative properties, then our caring about whether our actions exemplify normative properties may ultimately be deleterious to the promotion of what we care most about. We would then ought not to care about whether

our actions exemplify normative properties. It is, however, consistent to admit this and still think that there are such properties.

The thought that we ought not to care about whether things exemplify normative properties is not self-contradictory. It is consistent so long as we accept its implications–for example, the implication that, if it's true that we ought not to care about whether our actions exemplify normative properties, then we ought not to care that we ought not to care about whether our actions exemplify normative properties.

2.4 Irreducibility

Normative truths and domain-specificity

I mentioned in §2.3 that Scanlon thinks that our practical deliberation does not give rise to any live, non-trivial question as to why our reasons are normative for us. These questions, Scanlon suggested, are suitably answered by the consideration that our reasons *are reasons for us*. Nevertheless, Scanlon does admit that there is a meaningful, non-trivial question as to why reasons are normative:

> I believe that there can be meaningful external questions about a domain. But these must be questions about whether the implications or presuppositions of statements internal to the domain are fulfilled. (Scanlon 2014, p. 19)

Meaningful questions about why reasons are normative for us do not arise when agents are deciding what to do. When we are deciding what to do, we are asking the question "what reasons do I have?" Considering questions about what reasons we have takes place from a standpoint of "internal" reasoning, and it is, for Scanlon, the nature of "the point of view of

internal reasoning that is primary in an investigation of reasons and normativity. From this point of view the question of how reasons 'get a grip on one' properly disappears" (ibid., p. 14). The non-trivial question as to why reasons are normative for us surfaces when we consider reasons from a standpoint *external to* the standpoint of a practical deliberator.

In saying that there can be meaningful external questions about a domain, I take Scanlon to be saying, for example, that error theory is coherent. Error theory about normative claims, again, is the theory that any positive ascriptions of normative properties are false.[17] In making this claim, the error theorist is answering "questions about whether the implications or presuppositions of statements internal to the domain are fulfilled." It *makes sense*, for Scanlon, for error theorists to say that there are no normative truths. But in giving their error-theoretic arguments, error theorists are making a crucial mistake: applying scientific standards for specifying a body of truths to the normative domain in which these standards are not appropriate.

It's helpful to compare Scanlon's thoughts about internal and external questions about normativity to the foundational distinction I drew in §1 between what reasons are and what the reasons are. Scanlon is saying that the kinds of questions we consider when engaging in practical reasoning are questions about what the reasons are. The question what reasons are is a question external to practical deliberation: our asking this question entails that we are not engaging in practical reasoning. So when we are reasoning about my work in this dissertation, which deals exclusively with questions about what reasons are, we are not

[17]Error theorists don't think that *all* normative claims are false. For instance, they usually think that it's true that global warming is not bad. Not because they think that global warming is good, of course, but because they do not think anything is bad and therefore think that it's correct to deny that global warming instantiates any normative properties such as goodness or badness.

engaging in practical reasoning.

This phenomenon whereby engaging in practical reasoning "screens off" the meaningfulness of questions external to reasoning of this kind is a feature of other domains of reasoning as well. Scanlon writes,

> I believe that the way of thinking about these matters that makes most sense is a view that does not privilege science but takes as basic a range of domains, including mathematics, science, and moral and practical reasoning. It holds that statements within all of these domains are capable of truth and falsity, and that the truth values of statements about one domain, insofar as they do not conflict with statements of some other domain, are properly settled by the standards of the domain that they are about. (ibid., p. 19)

The domain of mathematical reasoning carries with it a set of criteria for specifying the truths of mathematics. Committing oneself to these mathematical truths does not produce any ontological worries, since none of the mathematical truths conflict with truths specified by the criteria for specifying the truths of other domains. In contrast, Scanlon thinks that claims about the nature of poltergeists, demons and so on do conflict with truths about other domains—namely, truths about the scientific domain. Theories of poltergeists and demons make claims about magical or spiritual laws and so on, the existence of which conflicts with the body of truths entailed by scientific criteria for specifying the scientific truths. This gives us reason to reject any claims that would be true given the theory of poltergeists.

In case this is not fully clear, we can imagine that the theory of poltergeists states that poltergeists have powers of telekenesis. At least, this is part of what popular culture seems to think poltergeists would be like. However, our best psychological and physical theories entail that there are no telekenetic powers. So the truths entailed by the theory of poltergeists conflict with the truths entailed by scientific theories. We therefore have reason to reject

claims whose only justification is that they are entailed by the theory of poltergeists. Scanlon is not clear as to why he thinks we should go with the scientific truths rather than the purported truths about poltergeists in the case of this conflict, but I assume that my readers have their own views as to why science rather than poltergeist theory is the way to go here.

One worry is that Scanlon's approach licenses more legitimate domains of reasoning than we have reason to believe there are. Take a religious reasoning system R that specifies its own set of domain-specific criteria for determining the extent and distribution of the property of being sinful, but does not make claims that conflict with the domain of scientific truth because R entails that sin will only play a causal role in determining our position in an afterlife which is in principle causally isolated from our own universe. Since truths about sin are domain-specific to R and do not conflict with the truths of science, one might think Scanlon has to admit that there are truths about the distribution of sin. One response Scanlon could make is that, although R does not make truth claims that overreach into the domain of truth regulated by scientific criteria for truth-determination, the claims that R entails about the role of sin in determining punishment in the afterlife overreach by making claims that conflict with the domain of normative reasoning. Since we have more reason to go with the normative criteria for determining truth than with criteria for determining truth given by R, we ought to reject R.

This domain-specific view of truth that Scanlon is advancing here is not a pragmatist view, because it does not say that the sole criterion of truth is that truth's usefulness. Usefulness may be one criterion of truth in some domains, but it is not a criterion of truth in every domain and certainly not the only criterion. Nor, I think, does Scanlon's domain-specific view entail that "true" is ambiguous, but rather that the criteria for truth vary from

domain to domain. Scanlon's view does bear similarities to Hirsch's quantifier variance view (Hirsch 2011, pp. 68-95) and Putnam's conceptual relativism (Putnam 1987, p. 71).

Dodging metaphysics?

Scanlon, furthermore, is not claiming that *objects* exist only within a domain or relative to a domain. In particular, he is emphatically opposed to the claim that there are normative objects that exist in the normative domain but not in other domains. The use of the word "domain" may give this impression, since in first-order logic we talk about domains of discourse as the set of objects that bound variables take as values. In Scanlon's usage of "domain", however, a domain is not a distinct set of objects, but a distinct set of truths entailed by rules of that domain for specifying these truths. I take Scanlon to be saying that domains are logical models. As such, domains have a domain of discourse, a language of constants and predicates for forming sentences about objects in the domain, and a set of non-logical axioms that, in addition to the logical axioms given by the domain's logic, specify which statements in that language are true. Importantly, for Scanlon, any domain has the *same set of objects* for its domain of discourse. He writes,

> The normative domain, for example, is not a distinct realm of objects. Things in the natural world, such as persons and their actions, have normative properties, and most normative claims are claims about such things. Even in the case of arithmetic, although there are pure claims about numbers, there are also numerical claims about the physical world. So a domain is better understood in terms of the kind of claims it involves, and hence in terms of concepts that it deals with, such as number, set, physical object, reason, or morally right action. (Scanlon 2014, p. 19)

What is special about normative claims is thus not a matter of ontology

> in Quine's sense (the things quantified over), but rather of what Quine called "ideology" (the predicates employed). (ibid., p. 30)[18]

By refusing to postulate the existence of distinct normative entities, Scanlon takes himself to be avoiding any ontological commitment and therefore any philosophical concerns that one might have about it. However, although Scanlon claims that he can avoid ontological commitment and any problems associated with it, he is happy to admit that questions about the reasons relation are "metaphysical" questions. He writes, "the basic element of the normative domain is a relation, being a reason for, [which] can be seen as a claim about the metaphysics of the normative" (ibid., pp. 25-26). The idea here seems to be that ontology is a branch of metaphysics: the one that deals with the question what we should allow to be taken as values of bound variables. Metaphysics is more capacious, including not only these ontological questions, but questions about properties and relations as well. Since the basic element of the normative domain is a relation—the favoring relation—discussion of this relation falls under metaphysics. But since the normative domain entails no truth about normativity that conflicts with the truths entailed by scientific theories, Scanlon holds that there is nothing metaphysically objectionable about the ideology of normative properties and relations.

Scanlon's view here marks a significant break with the traditional understanding of what's at stake in postulating, as Scanlon does in his own way, the existence of distinct normative properties. Critics of the intuitionisms of Moore, Prichard,[19], Ewing[20] Ross[21]

[18]The seminal discussion of this in Quine is (Quine 1951b).

[19]For example, (Prichard 1912).

[20]For example, (Ewing 1939).

[21]For example, (Ross 2007).

et al have frequently charged intuitionists not with having a problematic ontology of abstract normative objects, but with introducing additional, metaphysically problematic properties.[22] I had mentioned in §1 that Moore and the other intuitionists held that goodness was a "sui generis" property. By this, the intuitionists did not simply mean that the goodness property was distinct from other properties; anyone who thinks that there is a goodness property thinks that. Moore meant that goodness was a unique *kind* of property. Goodness, for Moore, was a property of neither the "natural" (i.e. observable, scientific) nor the "metaphysical" (i.e. spiritual, theological) kind, but of some third kind, about whose nature Moore had little to say. Moore's view, then, was not only, to use a distinction from (Lewis 1973, p. 87), less *quant*itatively parsimonious than the analytic naturalist views that Moore rejected, but less *qual*itatively parsimonious as well. Since the analytic naturalists identified the goodness property with some other natural property, such as pleasure, whose existence Moore was happy to accept, Moore's claim that there was a goodness property that was not identical to any natural properties entailed that Moore's ontology was quantitatively more capacious by at least one property. Moore's view that the goodness property was sui generis also committed Moore to accepting an additional kind of property, which is to say that his view was qualitatively less parsimonious as well. Most philosophers accept qualitative parsimony to be a virtue of philosophical theories such that, ceterus paribus, a more qualitatively parsimonious theory is to be preferred over a less qualitatively parsimonious theory.

Scanlon's turn to Quine's theory of ontological commitment therefore does not answer

[22]The first discussion of these issues that I am aware of is at (Frankena 1963, pp. 86-7). The discussion was of course popularized by (Mackie 1977, pp. 38-42) and is now omnipresent.

the charge that Scanlon's theory, like Moore's, is not qualitatively parsimonious. Furthermore, the discussion of ideology doesn't seem relevant either, since the metaphysical worries surface not from the nature of the predicates that feature in normative theories, but from the properties and relations that these predicates denote.

Scanlon could answer the worries by applying Quine's views on the nature of properties to the case of the favoring relation. Quine was an extensionalist about properties: he believed that properties were nothing but classes, and that classes consisted of nothing but their members. Quine writes,

> Once classes are freed of any deceptive hint of tangibility, there is little reason to distinguish them from properties. It matters little whether we read "$x \exists y$" as "x is a member of the class y" or "x has the property y." If there is any difference between classes and properties, it is merely this: classes are the same when their members are the same, whereas it is not universally conceded that properties are the same when possessed by the same objects.... But classes may be thought of as properties if the latter notion is so qualified that properties become identical when their instances are identical. (Quine 1951a, p. 120)

If he were to apply extensionalism about properties to the case of the favoring relation, Scanlon could say that the favoring relation were nothing except some specified set of reasons-apt tuples (i.e. some specified set of fact-agent-action-circumstance tuples). Call this set Γ. Since nearly all parties in metaethics are fine with accepting that there are such things as classes, facts, agents, actions and circumstances, the extensionalist reduction of the favoring relation to the aforementioned set of quadruples would resolve any outstanding worries about the metaphysical implications of Scanlon's construal of the favoring relation.

However, Scanlon can't accept that the favoring relation is nothing more than a class of fact-agent-action-circumstance quadruples. The first reason is that he thinks that there are

"irreducibly normative facts about reasons" (Scanlon 2014, p. 3). If the favoring relation were nothing more than a class of quadruples Γ, then facts about something's being a reason would not be irreducible: they would just be facts about that fact's being an element in a quadruple in Γ.

The second reason is because Scanlon thinks that there is a sound open question argument about the favoring relation. Scanlon writes "I am quite willing to accept that 'being a reason for' is an unanalyzable, normative, hence non-natural relation" (Scanlon 1998, p. 18). He also writes, "[O]pen-question arguments show that neither claims about what counts as evidence nor claims about what count as reasons for action can be plausibly understood as claims about natural facts (ibid., p. 60).

The open question argument was first articulated by Moore.[23] Moore argued that our concept of goodness was simple because there was no analysis of the goodness concept in terms of simpler concepts. His argument for this latter claim was an inductive one. Moore claimed that, for all definitions Y, if there is an object x such that a "yes" answer to the question "object x satisfies definition Y, but is X good?" is informative, then Y is not a correct and complete analysis of goodness. Moore then applied this principle to a variety of different definitions of goodness that had been proposed in the literature at the time—that "good" meant "pleasure", or "utility" or "what we desire to desire"—and found that the "yes" answer was in fact informative in all these cases. This forms the basis of an inductive argument that no adequate analysis of the goodness concept in terms of simpler concepts was to be had. Because there is no adequate analysis of the goodness concept, Moore

[23] Sidgwick did claim that "ought" was primitive and irreducible, but Sidgwick did not draw these conclusions from open question considerations. Moore was the first to do that.

ultimately concluded that the goodness property was distinct from any other property that can be expressed by our other concepts.

To put it another way, the openness of so many questions about whether an entity with such-and-such a property is good provides defeasible evidence that the fact of any entity's being good is distinct from any fact that that entity has some other property. The evidence would be defeated in the event that someone asked such a question that lacked the openness. But no one has ever done this.[24] So the evidence in favor of the distinctness of goodness from natural properties, though defeasible, continues to favor the irreducibility of goodness.

Scanlon accepts that there is a sound open question argument that establishes the irreducibility of the favoring relation. For example, Scanlon would concur that "the quadruple consisting of Peter Singer, the fact that there is a drowning baby in front of Singer, Singer's action of saving the drowning baby, and Singer's circumstances is in the set Γ, but is the fact that there is a drowning baby in front of Singer a reason for Singer to save the drowning baby in those circumstances?" is an informative question that requests a non-trivial answer. Since he is also committed to the soundness of the open question argument, he should therefore conclude that the favoring relation is not identical to the set Γ, in other words, that it can't fully consist in this set and must be something over and above it.

In the next section, I give a detailed investigation and criticism of the use of the term "non-naturalism" in metaethics literature. This section does not directly contribute to the overall argument of this dissertation, and so readers pressed for time can skip it.

[24]Some people say that they have: Frank Jackson (Jackson 1998), Mark Schroeder (Schroeder 2007), David Lewis (Johnston 1989), Michael Smith (Smith 1994).

What "naturalism" doesn't mean

Scanlon's commitment to irreducibly normative facts makes him a kind of metaethicist that's known as a "non-naturalist". Non-naturalism is the view that normativity is "what it is and not another [natural] thing." If this is confusing, it's because the term "non-naturalism" is itself a confusing term. In my view, it's also quite unhelpful and probably should be exorcised from contemporary metaethical discourse.[25]

First, it is misleading. In suggesting that normative properties are not natural properties, it suggests that error theorists are non-naturalists, since error theorists think that there are no normative properties and therefore that no natural properties are normative properties. They should therefore happy to claim that normative properties are "non-natural" in the sense that normative properties are not natural properties. But error theorists have been among the most influential and incisive critics of those theories that have fallen under the "non-naturalist" label.

Secondly, the term comes with a great deal of historical baggage. We owe the origin of the term to Moore's influential *Principia Ethica* which, despite its role in drawing the distinctions that are responsible for the creation of metaethics as a discipline, features a lot of arguments that, from our contemporary vantage point, strike us as immature and unsound. No one thinks that Moore is right about everything he claims about metaethics.[26]

Thirdly, labeling a theory as "non-naturalist" tends to produces in many of us a feeling of contempt and dismissal towards that theory that constrains our ability to accept or reject

[25] I'm not suggesting that it should be exorcised from general philosophical discourse, but just from metaethical discourse.

[26] Including Moore, who rejects his own claims in chapter 1 of *Principia* at (Schilpp 1968, pp. 581-2).

the theory on the basis of its merits.[27]

Most problematically, "non-naturalism" about normative properties is a denial of naturalism about normative properties, but there is no agreement in metaethics on what naturalism about these properties is supposed to be. Consequently, there is no agreement in metaethics on what non-naturalism about these properties is supposed to be.

In the remainder of this section on irreducibility, I will discuss what metaethicists have thought they meant by natural property, give some reasons why their suggestions are misleading, and then offer my own alternative formulation of the issue metaethicists are talking about when they talk about natural properties.

Sparse properties

One reasonable approach to resolving questions about what naturalism about normative properties is would be to defer to professional metaphysicians. Although this would normally be a reasonable thing to do, it's not a good strategy for uncovering what metaethicists mean by "naturalism".

When metaphysicians talk about natural properties, they tend to be talking about "sparse properties": the fundamental constituent properties of the universe. These are contrasted with "abundant" properties: the gruesome properties as well as other sorts of disjunctive properties. Lewis describes perfectly natural properties in this way:

> Sharing of [the perfectly natural properties] makes for qualitative similarity, they carve at the joints, they are intrinsic, they are highly specific, the sets of their instances are not ipso facto entirely miscellaneous, there are only just

[27] This knee-jerk skepticism about "non-naturalist" views is not unique to philosophers. My students have it, and my mom and therapists do too.

> enough of them to characterize things completely and without redundancy. (Lewis 1986, p. 86)

The sparse property interpretation of natural properties as metaethicists talk about them is mistaken. For example, Frank Jackson argues that normative properties are disjunctive properties (Jackson 1998). But everyone agrees that Jackson is a normative naturalist. Furthermore, on Moore's view, the goodness property is a simple, intrinsic property that's needed in order to characterize things completely. If this view of goodness were correct, goodness would be a sparse property and therefore a natural property on Lewis' usage of the term "natural". However, Moore is the paradigmatic normative non-naturalist. Whether a theory entails that normative properties are among the abundant properties or the sparse properties is therefore neither necessary nor sufficient for that theory's being a natural or non-natural theory of normativity, as metaethicists talk use the terms.

Ostensive definitions

On another characterization of natural property, natural properties are ones that are "similar to" the properties of a certain set of things that we point to. Loosely, on the ostensive approach, we point to some things, define a class of things Γ as the class of all the things that are "like" the things we pointed to, then say that the natural properties are the properties possessed by one or more of the elements in Γ.

Jackson proposes an ostensive characterization of naturalism in a discussion of naturalism in the philosophy of mind. He writes,

> Physicalists can give an ostensive definition of what they mean by physical properties and relations by pointing to some exemplars of non-sentient

> objects—tables, chairs, mountains, and the like—and then say that by physical properties and relations, they mean the kinds of properties and relations needed to give a complete account of things like them. Their clearly non-trivial claim is then that the kinds of properties and relations needed to account for the exemplars of the non-sentient are enough to account for everything. (ibid., p. 7)

In the context of the philosophy of mind, it's essential to the ostensive strategy for specifying the natural properties that the objects one points to in the ostension procedure be non-sentient. Jackson is not a naturalist about consciousness, and so if conscious objects were implicated in the ostension procedure, the ostension procedure would include consciousness properties among the list of natural properties. Jackson does admit that the truth of panpsychism—the view that all objects have consciousness properties—would frustrate the ostensive strategy for demarcating the natural from the non-natural properties. However, Jackson is not bothered by this, since he thinks panpsychism is such a fringe view that one does not need to take it seriously.

The worries about panpsychism mirror worries that one may have about giving an ostensive characterization of non-natural properties with respect to the normative sphere. "Pannormativism", unlike panpsychism, is not so fringe that we can dismiss it out of hand.[28] It's quite plausible that all, or at least the vast majority, of objects that we could in fact indicate in an ostensive procedure have normative properties. There are good and bad tables and chairs, and mountains can be ugly or beautiful. Another way to put the worry is that, while we do have a way to indicate, ostensively, which objects are non-sentient, we do not have a way to indicate, ostensively, which objects are non-normative, since exemplifications of normative properties are much more pervasive and ubiquitous than exemplifications of

[28] I also think that panpsychism is worth taking more seriously than Jackson takes it, especially given the recent resurgence of interest in panpsychism due to Chalmers' influence. But I set this quibble aside here.

consciousness properties.

Disciplinary criteria

The proposal that one finds the most in metaethics discussion of conditions for natural property is that natural properties are the properties denoted by the concepts constituting natural scientific theories. Shafer-Landau's account is a good example of this. He suggests that natural properties are whatever properties are objects of study by the natural sciences, where natural sciences are understood to be those "sciences whose fundamental principles are discoverable a posteriori, through reliance primarily on empirical evidence" (Shafer-Landau 2003, p. 212). Shafer-Landau claims that his characterization "is true to all instances of avowed ethical naturalism that [he knows] of, and does seem to capture what many have thought to be essential to the classification" (ibid., p. 212).

Shafer-Landau's optimism is misplaced. The main objection I have to this is that that sciences all accept, as fundamental principles, basic metaphysical necessities, at least some of which are not discoverable a posteriori, even in principle. As a result, the definition entails that no domains of inquiry are natural sciences.

Take for instance the right-to-left direction of the following infamous principle of metaphysics:

Leibniz's Law: $\forall x \forall y[x \approx y \leftrightarrow (P(x) \leftrightarrow P(y))]$.

This is a principle that underlays every natural science. When physicists see the track in the electron cloud chamber, they postulate the existence of one electron, not seven indiscernible, consistently co-located entities with one-seventh of the mass and charge of a typ-

ical electron. But this principle is not discoverable a posteriori, even in principle. There is nothing that we could do to test whether the entity that caused the track in the cloud chamber is one electron or seven consistently co-located entities with one-seventh of the mass and charge of a typical electron. There are also the logical laws, such as $\forall x[x \approx x]$, which feature in scientific theories but are not discoverable a posteriori even in principle.

Shafer-Landau's disciplinary criteria of natural property to work, it would need to be subsidized with some kind of further limitation on which fundamental principles are the ones that need to be discoverable a posteriori in order for the theory in which the principles feature to qualify as a natural science. But this is to supplement the disciplinary approach with an epistemic one. The most developed epistemic approach to characterizing natural properties in metaethics is given by David Copp, and I'll consider his approach in the next section.

Epistemic criteria

David Copp offers the most detailed and sophisticated characterization, epistemic or not, of natural properties in the metaethics literature. According to Copp, natural properties are just *natural** properties, where "natural*" is defined in the following way:

> *Copp's Criteria*: A property *N* is natural* if and only if (a) it is possible for *N* to be instantiated and (b) there are propositions about the instantiation of *N* that are both synthetic and possibly true, and, (c) no such proposition is strongly a priori. (Copp 2003, p. 183)

We can safely ignore the first (a) criterion. I'm actually not sure why Copp includes it, as he never discusses this criterion in his work. Furthermore, it's entailed by the Principle

of Instantiation which is accepted by most theorists of properties. The Principle of Instantiation states that there are no uninstantiated properties or, equivalently, that every property is instantiated (Armstrong 1989, p. 75). Now, for any property *N*, if *N* is instantiated, then it is possible for *N* to be instantiated. If this is true for properties in general, then it will be true a forteriori for properties of a certain kind, such as the natural* kind of properties. So we can just leave (a) to the side.

Copp's notion in (c) of a *strongly a priori* proposition expands Hartry Field's analysis of an a priori proposition (Field 2001, pp. 117-118). Field specifies that an a priori proposition is a proposition that is both *weakly a priori* and *empirically indefeasible*. A weakly a priori proposition is a proposition that can be reasonably believed without empirical evidence. An empirically indefeasible proposition is a proposition that admits no empirical evidence against it. Strong apriority on Copp's definition is a weakly a priori proposition that is empirically indefeasible *for an "ideal thinker"*: "a thinker with no psychological weaknesses, with no computational limitations, and with a full conceptual repertoire" (Copp 2003, p. 190). Copp adds this last condition in order to avoid Phillip Kitcher's concern that our beliefs about certain mathematical properties—say, "being such that Fermat's last theorem is true", would be defeated by possible testimonial evidence from mathematicians that claim that they have discovered a counterexample (Kitcher 1983).

The (b) condition is meant to constrain the (c) condition only to those propositions about the instantiation of a property that are synthetic. This is needed because propositions about the instantiation of natural properties are sometimes analytic and therefore strongly a priori (on standard readings of analyticity). The proposition "quarks are physical" is a proposition about a certain property that quarks have: the property of being physical. It will come out

analytic, since the truth of this proposition follows from the meaning of "quark" alone, and will therefore come out strongly a priori. There would then be a strongly a priori proposition about the property of being physical, which would mean that the property of being physical would fail to satisfy (c) and therefore not qualify as a natural property. That would be horrible. So we need to restrict the strongly a priori propositions that interest us to non-analytic propositions about the instantiation of properties in order for "being physical" and many other natural properties to avoid coming out non-natural on Copp's reading.

Unfortunately, the conditions Copp offers are not sufficient for a property's being natural. Copp makes the following claim: "For example, people who accept the story of creation in Genesis as the literal truth hold that God caused the world to exist. It would muddy the water to take their view as naturalistic" (Copp 2003, p. 184). However, Copp's criteria actually serve to categorize this view as naturalistic. Here is why.

Take the proposition "the universe is God-created", by which I mean created by the God of the Abrahamic traditions, with the standard properties of being all-powerful, all-good and so on. "The universe is God-created" satisfies condition (a) of Copp's criteria, since it's possible that the universe instantiates the property of being created by God.[29]

Condition (b) is satisfied, since, on orthodox theistic views, claims that God created the universe are both true (therefore possibly true) and also synthetic rather than analytic.[30] Even on Anselm's rationalistic picture of divine knowlege, it is analytic that God exists, but

[29]This would be false if the Abrahamic God's existence were metaphysically impossible. I take it that most philosophers take the Abrahamic God not to exist, but not that God's existence is impossible. In fact, one is more likely to get closest to such claims not in atheist philosophers of religion but in orthodox theologians such as Duns Scotus, who claims that God's existence is either necessary or impossible.

[30]I am not claiming here that orthodox theistic views entail that there are no analytic truths about divine reality. I'm only asserting that they assert that the truth or falsity of claims about God's having created the world are synthetic.

not analytic that God *created the universe*. Anselm would have known that among the list of medieval Christian dogmas is the dogma that God's creation of the universe was free. In other words, God was free *not* to have created the world. If knowledge of God's nature involves knowledge that God is free to create the world or not, then we would need an additional piece of information to conclude which of these divine actions God chose. One would need to supplement one's concept of God with the further empirical (non-analytic) fact that *there is a world* in order to be justified in one's belief that God created the world.

Condition (c) is satisfied, since any claim that God created the universe is empirically defeasible and therefore not strongly a priori on Field's definition of strong apriority. Suppose we were to look around the world and see an incomprehensible amount of unmerited suffering. This would constitute empirical evidence that the world was not created by the all-powerful, all-good deity that Abrahamic theists believe exists.

"God-created" then, satisfies all three of Copp's criteria and would then qualify as a natural property. But it's not, by Copp's own admission. So Copp's criteria are insufficient for establishing that a property is natural.

Here is another, more relevant counterexample to the sufficiency of Copp's Criteria as criteria for a property's being natural. Copp's criteria also qualify "prohibited by God" as naturalistic. But it's not.

On divine command theories, what it is for an action to be wrong is for that action to be prohibited by God or some other specified divine entity. However, on most divine command theories, the set of divine prohibitions is either revealed through some miraculous event like the Christ event (as John Calvin asserts), or such that at least some of the prohibitions are knowable in virtue of their constituting part of the natural law (as Aquinas asserts). In other

words, divine command theorists who believe in God[31] don't typically think that we have a priori knowledge of the inner workings of God's mind that can tell us what actions God prohibits. In order for us to know anything about what actions are prohibited and therefore what actions are wrong, God either has to manipulate the world through some miraculous event (giving angels natural causal powers, writing the laws in stone tablets, etc) to make known the set of prohibitions, or God has to have designed our intellects in such a way that we can make good inferences into the nature of the divine mind from the fact that the world that God created has the features that it does. But the epistemic conclusions that one would draw in either of these ways would be synthetic (empirical). They would also not be strongly a priori, even for an ideal thinker. This is because they would be empirically defeasible: if we discovered that the golden plates were carved not by Moroni but by John Smith, or that the world has different features than it in fact did, then we could be rationally required to rethink our conclusions about what actions God is prohibiting. "Prohibited by God" would therefore come out as a natural property, which, again, would "muddy the waters" as to what ought to count as naturalistic. No one in metaethics thinks that, if the property of being wrong just were the property of being prohibited by God, then wrongness would be a naturalistic property. This goes to show that Copp's Criteria not only fail to give an account of natural property simpliciter, but they also fail to give an account of natural property that respects the usage of the term "naturalism" in metaethics.[32]

[31] You could be a divine command theorist but fail to believe in God, in which case we'd have the "God is dead; everything is permissible" situation that Dostoevsky writes so eloquently about.

[32] With heady terms like "naturalism", with many different uses across myriad disciplines (philosophical and not) one may reasonably wonder who gets to say what counts as natural and what does not. Someone, for instance, might state that divine properties are perfectly natural and, in cases of disagreement such as these, one can reasonably ask "Who's right: this other person or me?" In my estimation, both are right–not because this would be a case of faultless disagreement, but because "naturalism" is multiply ambiguous and admits

Jettisoning "naturalism"

These previous reflections have aimed to show that there is no adequate characterization of what it is for a property to be natural as metaethicists use the term. In my view, part of why we have had trouble giving a characterization of natural property in metaethics is because metaethicists themselves are guilty of inconsistent usage. For instance, take Brink, a paradigmatic Cornell realist and a naturalist. Brink is a "non-reductive" realist: he thinks that normative properties are constituted by, but not identical to, the other natural properties that feature in the natural and social sciences. Shafer-Landau also thinks that normative properties are constituted by, but not identical to, the other natural properties that feature in the natural and social sciences. Yet Shafer-Landau is considered to be a non-naturalist even though Brink is considered to be a naturalist.

This general inconsistency and confusion among metaethicists in their use of the terms "natural" and "non-natural" is largely due to the terminological idiosyncrasies in Moore's original presentation of his variety of moral intuitionism. Moore himself gave four distinct characterizations of natural properties in his *Principia*: as intrinsic properties, as properties that exist in time, as properties that can be objects of perception, and as properties that can exist by themselves. None of these characterizations make any sense,[33] so it stands to reason why philosophers to this day remain confused about what non-naturalism about normative properties is supposed to be.

---

of a variety of definitions and possible precisifications. There is no one single definition of "naturalism" that can make sense of uses of the term across the many disciplines that make use of it. Indeed, if I am right, there is no one single definition of "naturalism" that can make sense of uses of the term even within the single discipline of metaethics. It is for this reason that this dissertation uses the technical term "irreducibility", which I characterize in the next section, rather than the multiply ambiguous term "non-naturalism".

[33]For more on why, see (Dreier 2006).

As I said at the beginning of this discussion of "naturalism" and "non-naturalism", I think that the terms "natural" and "non-natural" carry so much historical baggage that we should jettison them from our lexicon. Philosophers are happy to jettison confused and ambiguous folk terms in order to achieve philosophical precision, and it seems to me that the same principle can motivate jettisoning of confused and ambiguous philosophical terms as well. Shafer-Landau is right, I think, in suggesting that "we need to ask just how much of the [naturalism/non-naturalism] debate has become a matter of taxonomic bookkeeping, and how much of philosophical significance is really at stake. What matters is not what we call a view, but whether the view, whatever its denomination, can solve the central philosophical problems that have generated the division between naturalists and non-naturalists" (Shafer-Landau 2003, p. 62).

Furthermore, to the extent that the terms "naturalism" and "non-naturalism" in metaethics do make sense, they most likely refer to distinct views in the epistemology of normativity rather than the metaphysics of normativity. This is problematic, because the natural/non-natural distinction is almost exclusively cached out as a distinction between positions in the metaphysics of normativity. Jettisoning these terms and replacing them with two kinds of distinctions—a distinction between positions in the metaphysics of normativity and a distinction between positions in the epistemology of normativity—will help clarify the lines of debate and help metaethicists make more rapid progress.

Characterizing irreducibility

In this subsection, I will give my own characterization, not of naturalism in metaethics, whatever that is supposed to be, but of a view that I take to have clear philosophical significance for the metaphysics of normative properties. That view is *irreduciblism* or *irreducibility*. Irreducibility about the favoring relation, as I stated at the very beginning of this chapter, is the view that the favoring relation is *not* normatively masked. Reducibility about the favoring relation, in contrast, is the view that the favoring relation *is* normatively masked. In order to explain what irreducibility and reducibility are, then, I have to explain normative masking.

Normative masking

Here is the idea. You take all our non-demonstrative, n-ary predicate concepts and write them all down in a list, along with their reference conditions. The list of predicate concepts will include concepts like squareness, being tall, being morally good, and being a reason for an action for an agent in a circumstance. They will not include the concept of a table or Socrates, since these are not predicate concepts.[34] The reference conditions for a predicate concept will be those conditions that must be satisfied by a property in order for the predicate concept to refer to it. Then add in all the predicate concepts that we would need to construct all the true empirical theories, whatever those happen to be, along with their reference conditions as well.

Now, divide these concept/reference-condition pairs into two sets: the set A consisting of the normative concepts; and the set B consisting of (a) the non-normative concepts plus

[34] I assume, along with Scanlon, that there are no normative objects, but only normative properties.

(b) the non-normative predicate concepts needed to express all true empirical theories. Each of these pairs of concept and reference conditions will tell us what property or properties fall under the concept in a given world. In some cases, the reference will be the empty set, such as the concept pair consisting of the concept phlogiston and its reference condition. In other cases, the reference will be non-empty, such as the concept "being tall".

Take the set of all the properties denoted in a world w by one or more of the concept/reference condition pairs in A and call that $R_w(A)$. Take the set of all the properties denoted by one of more of the concept/reference-condition pairs in B and call the $R_w(B)$. Some normative property P is *normatively masked* in w just if it is a non-empty member of both $R_w(A)$ and $R_w(B)$. One is a reduciblist about P in w iff one holds that P is instantiated (exists) in w and is normatively masked, and one is an irreduciblist about P in w iff one holds that P is instantiated (exists) in w and is not normatively masked. In what follows, I will just assume that the world in question is ours.

The "non-demonstrative" clause is important. Take the demonstrative concept "what Jeff just denoted". This probably qualifies as a non-normative concept and as such would end up in one of the pairs in B. However, if Jeff just denoted the property of being morally right, and "what Jeff just denoted" ended up in B in virtue of its being a non-normative concept, then $R(B)$ then includes moral rightness, moral rightness ends up being normatively masked, and anyone who thinks that moral rightness is a property would qualify as a reduciblist. Since I'm trying to give a procedure for distinguishing reduciblists from irreduciblists, I can't allow this. So, as a response, I just exclude these sorts of demonstrative concepts from my account.

The "true empirical theories" clause is also important. Suppose that our current empiri-

cal theories are false and we have not yet done the empirical investigation needed to acquire the concepts we would need to construct the set of true empirical theories, whatever those may happen to turn out to be. Now we do that empirical investigation and gain the corresponding concepts. Some future philosopher Jones then says that moral goodness denotes some descriptive part of the world that we can't currently refer to in a non-demonstrative fashion. If the set of concept/reference-condition pairs in B do not include these new concepts, then there would be no property in $R(B)$ that C denotes, and Smith would be an irreduciblist about moral goodness. But we should not consider Jones to be an irreduciblist, since he is identifying moral goodness with a property that falls under a non-normative concept.

We can't actually enumerate all the concepts in A and B. In particular, my account assures that we cannot do so, at least not yet, if we assume (reasonably, I think) that we don't yet possess all the concepts needed to express all true empirical theories. Fortunately, I don't need us to be able to enumerate all such concepts. All I need is for us, in principle, to be able to distinguish between the normative concepts and the non-normative ones. By a "normative" concept, I mean the broadest class of concepts that directly express any content about how the world ought to be or could ought to be, and by a non-normative concept I mean the broadest class of concepts and property that do not directly express any content about how the world ought to be or could ought to be. I am, however, not providing a procedure as to how to demarcate the normative concepts from the non-normative concepts. I simply assume that there is such a procedure.[35]

[35]Many have been proposed. Searle, for instance, suggests that the distinction can be cashed out in terms of direction of fit (Searle 1989, p. 91). Normative concepts, Searle suggests, have a "mind-to-world" direction of fit, while non-normative concepts have a "world-to-mind" direction of fit. Also see (Schroeter and Schroeter

All I am doing here is providing a formal characterization of what is at stake in debates between what I'm calling irreduciblism and reduciblism about normative properties. In general, these debates will map onto what has traditionally been understood to be the debate between non-naturalism and naturalism about normative properties. However, as I have said, I find these terms misleading and think that they have confused the lines of the debate so that they no longer make much sense. I offer the characterization I offer here in an attempt to make these lines sharper.

If the normative/non-normative concept distinction is hopelessly confused, as Jackson worries (Jackson 1998, p. 121), then my proposed characterization of what's at stake between irreduciblism and reduciblism about normative properties is a non-starter. However, I think that this is a virtue of my approach. If the normative/non-normative concept distinction turns out to be a hopeless confusion, then we should not be surprised to find that various other metaethical theses that rely on such a distinction are also a hopeless confusion. Furthermore, I have already tried to argue that the naturalism/non-naturalism distinction is already at least a confusion, if not a hopeless one, so if the distinction between normative and non-normative concepts also turns out to be confused, nothing at least will have been lost by adopting my approach.[36]

2014) for a review of various other attempts to cache out the distinction, as well as Schroeter and Schroeter's own novel theory.

[36]Lastly, there is a position in logical space according to which we do not have any normative concepts. On this view, the normative/non-normative concept distinction is coherent, but we just happen not to have any normative concepts. I'm construing the normative properties as properties that stand under normative concepts. Since, on the no-normative-concept view, there are no normative concepts for properties to stand under, such a person would count neither as a reduciblist nor as an irreduciblist on my characterization of these views. However, I think that this is the correct result. The hypothetical defender of the no-normative-concept position that I just described must admit that they have no conception of what it might mean for a property to be a normative property, and so they would not be in a position to offer substantive theories about the metaphysics of normative properties. Reduciblism and irreduciblism are substantive theories of the metaphysics of normative properties, so defenders of the no-normative-concept view therefore cannot

Normative masking in action

To see my account of irreduciblism at work, we can apply it to the example of reductive hedonism about goodness: the view that goodness is numerically identical to pleasure and the absence of pain. For the reductive hedonist, the goodness property stands under two distinct concepts: the concept "good" (a normative concept), and the concept "pleasure and the absence of pain" (which, we assume, is a non-normative concept). If reductive hedonism about goodness is true, then goodness is normatively masked, since it is a normative property but it stands under the non-normative concept of pleasure and the absence of pain.

Similarly, Schroeder is a reductivist or reduciblist about the favoring relation. For Schroeder,

> For some fact *R* to *favor X* to do action *A* in circumstances *C* is for there to be some state of affairs *p* such that *X* has a desire whose object is *p*, and the truth of *R* is part of what explains why *X*'s doing *A* promotes *p* in *C*. (modified from (Schroeder 2007, p. 59))

For Schroeder, the favoring relation *F(R, X, A, C)* stands under two concepts: the concept of R's *favoring* X's A-ing in C (a normative concept); and the concept of R's *partially explaining* why A satisfies a desire of X in C (a non-normative concept). The favoring relation is therefore normatively masked, since it is a normative property but stands under a non-normative concept: that of a partial explanation of the promotion of a desire of an agent in a circumstance. The irreduciblist will accept that there is such a thing as a partial explanation of why an agent's doing some action would satisfy one of her desires in some

accept either such theory. In terms of their semantics, I suspect that they would have to come out either as error theorists or as quietists about normative discourse. I suspect that a lot of philosophers working outside metaethics defend positions that do ultimately commit them to the no-normative-concept view. However, since no one, to my knowledge, defends this position in print, I do not treat it more extensively in the body of this dissertation.

circumstance. The irreduciblist simply denies that this is what it is for a reason to favor some action of an agent in a circumstance, since the irreduciblist about the favoring relation simply claims that there is no non-normative concept that the favoring relation stands under.

Virtues of the normative masking approach

The reducibility/irreducibility issue is philosophically important and philosophically helpful. It is fundamentally the issue of what normative properties are. If Schroeder's theory reduciblist theory of reasons is true, then what it is to be a reason is to be a partial explanation of why an agent's doing some action would satisfy one of her desires in some circumstance.

Casting the debate between reducibility and irreducibility also helps to isolate normative metaphysics from normative epistemology. Normative epistemology is really important; however, it is not normative metaphysics and should be understood as a separate area of metaethical inquiry. In particular, Brink and Shafer-Landau are both irreduciblists on my account, but they differ on the question of moral epistemology: Shafer-Landau thinks that ethics is an a priori discipline, while Brink thinks that it is a matter of a posteriori investigation into which irreducible properties causally regulate our use of normative terms. Different reduciblists also take distinct approaches to moral metaphysics. Jackson, for instance, is a reduciblist who thinks that ethics is an a priori discipline; while Boyd thinks that moral goodness, for instance, reduces to those first-order systems that promote and maintain human homeostatic equilibria, and thus that we need empirical investigation to determine what systems promote and maintain human homeostatic equilibria. There are, therefore, at least four major positions about normative metaphysics and normative epistemology that

metaethicists should consider:

Metaphysics↓/Epistemology→	apriority	aposteriority
irreducibility	e.g. Shafer-Landau	e.g. Brink
reducibility	e.g. Jackson	e.g. Boyd

I have argued that the non-naturalism/naturalism distinction in metaethics is confused and possibly incoherent. If the distinction does turn out to be coherent, I suspect that the distinction would have to amount to the distinction between, on the one hand, the a priori irreduciblist position indicated in the typology chart, and, on the other, the other three positions indicated in the chart.[37] This is an unhelpful way to divide the logical space expressed in the chart, as it ignores important differences between the three views that "naturalism" lumps together. My approach to understanding the logical space helps both to clarify the lines of metaethical debate and to underscore that there are more positions in logical space than the non-naturalism/naturalism distinction allows.

Virtues of irreduciblism

Why be an irreduciblist? In this section, I will give four arguments in irreduciblism's favor.

First is the infamous open question argument, first articulated by Moore.[38] Although, on my view, Moore made important and hugely influential contributions to ethical theory,

[37] I actually expect that "naturalism" is ambiguous as to whether the a posteriori irreduciblist position counts as naturalism. As I will discuss in §4, I think that Brink's view is not clearly a naturalist view, despite its commitment to an a posteriori epistemology of normativity.

[38] Sidgwick did claim that "ought" was primitive and irreducible, but Sidgwick did not draw these conclusions from open question considerations. Moore was the first to do that.

Moore's version of the open question argument has issues. One major issue is its implicit reliance on something like the following principle:

> *Primitivism-Entails-Irreducibility*: For all terms that denote properties, if a term expresses a primitive concept, then the property that that term denotes is an irreducible property.

Moore made implicit appeal to this principle in the open question argument. However, Primitivism-Entails-Irreducibility admits of many counterexamples. The concept "I" is primitive and refers (in my mouth) to me. But the me that it refers to is whatever I am: biological individual, set of interconnected memories or what have you. Our "heat" concept is a primitive concept and refers to heat. But the heat that it refers to is molecular motion. Our "time" concept is primitive and refers to time. But the time that it refers to is a component of Minkowski spacetime. Me, heat, and time are not irreducible, since they (we) reduce respectively to a biological individual, molecular motion, and a component of Minkowski spacetime despite the primitivity of the concepts that denote them (us).

Although these are legitimate concerns about Moore's argument, to dismiss Moore's views on the basis of one's dismissal of this argument would be to too quickly dismiss these views. Kit Fine (Fine 2002) claims that Moore's real argument for the irreduciblist nature of goodness was not the oft-cited material from Chapter 1 of *Principia Ethica*, but rather a lesser-quoted article of Moore's called "The Conception of Intrinsic Value". The insight expressed there, according to Fine, is not that "good" does not admit of an adequate analysis, but that, for any purported *real definition* of goodness (say, of promoting pleasure over pain), we have a strong intuition that something's being good consists in something more than the purported real definition. In other words, even if it were a conceptual truth

that something is good iff it promotes pleasure over pain, there is still the question why this would be the case. If the answer to this question were "something's being good consists in nothing over and above its promoting pleasure over pain", then it would be reasonable to suggest that the conceptual truth that something is good iff it promotes pleasure over pain holds because goodness would then just be the promotion of pleasure over pain. However, we have an intuition that goodness consists of something more than the promotion of pleasure over pain. So, even if it were a conceptual truth that something is good iff it promotes pleasure over pain, this itself is not sufficient to establish that goodness just is the promotion of pleasure over pain. Moore's irreduciblist conclusions could then still be vindicated even if someone does discover that, necessarily, for any x, x is good iff x promotes pleasure over pain.

I think that Fine is right about this. I think that Fine is right that we have a metaphysical intuition that the goodness of something consists in something over and above this thing's promotion of pleasure over pain, or any other property that this thing possesses. The truth of this has important ramifications on our understanding of the relation between evidence for irreducibility and the open question argument. I suggested earlier that, if there were *no* nominal definition P of goodness such that there were a non-informative question of the sort "x is P, but is x good?", then this provides defeasible warrant for the irreduciblist conclusion. It's defeasible, of course, because we could discover a posteriori that goodness was coextensive with the promotion of pleasure over pain, which raises the issue of reducibility. However, if someone were to propose "promotion of pleasure over pain" as a real definition of goodness, the irreduciblist can call back on the Finean intuition that something's being good consists in something over and above its promoting pleasure over

pain to defend irreduciblism even against the failure of open question phenomena.

A third argument is an argument from the existence of commitments and akrasia. "Commitment" is an umbrella term that references a variety of mental states and linguistic entities: intentions, promises, convictions, to name a few. What they have in common is that they set standards that we are *committed* to upholding. That's why they are called commitments. All of us can think of a time when we have had a commitment—to a deadline, a loved one, a student—and failed to live up to it. In such a case, we experienced what is sometimes called *akrasia*: weakness of will, a failure to live up to what one has willed to do. In fact, it is possible for someone to be committed to something even if they never ever live up to the standard. One stark example is athletic competition. For every Olympic sport, many people are committed to winning it, yet only one will actually live up to this commitment. The fact that one can have a commitment to some state of affairs X obtaining and yet never live up to this standard indicates that a commitment to X cannot be simply a disposition to X. Commitments, then, are a distinct kind of mental state than dispositions. But commitments are normative. So commitments present an example of a normative state that fails to reduce to non-normative states such as dispositions.

A fourth argument for irreduciblism is given by Brink (Brink 1989, pp. 158-159). Brink argues that normative terms are causally regulated by multiply realizable normative properties. What this means, for example, is that, in our world, our term "injustice" will refer to a certain concatenation of economic and social properties that realize the property of injustice. However, there will be other worlds in which these specific economic and social properties will not realize the property of injustice, and worlds in which other economic and social properties do realize the property of injustice. For example, the economic condition

in which Nestle owns all the water rights in large swaths of Kenya realizes injustice in our world, but in worlds in which people in Kenya didn't need water or had sufficient financial assets to pay Nestle for water, this particular economic condition perhaps would not realize the property of injustice. Furthermore, the economic condition of harvesting seaweed to sell to consumers does not realize the property of injustice in our world, but they might in worlds in which seaweed were sentient. Now, we could say that the property that causally regulates "injustice" was some disjunctive property, in which case we could say that the property that causally regulates "injustice" just was the property "monopolizing of water access or harvesting sentient seaweed or ... or ...." However, there are no disjunctive properties. So we cannot say that this disjunctive property is the property that uniquely causally regulates our use of "injustice". Ditto for the other normative properties. I discuss this argument for irreducibility in more detail in §4.

Concluding Irreducibility

In this section §2.5, I have discussed Scanlon's approach to normative metaphysics, argued that this approach is insufficient, discussed how this approach maps onto a position that metaethicists call "non-naturalism", argued that the use of this term is insufficiently precise to delineate an actual metaethical position, offered my own characterization of positions on a major issue in normative metaphysics that I think better delineates what these positions are supposed to be, and then discussed virtues both of this characterization and of irreducibility in general. In the next section, I will move on to motivating and characterizing the reasons fundamentalist's commitment to supervenience.

2.5 Supervenience

Reasons fundamentalists agree with basically everyone that normative reality "covaries" with non-normative reality: the world's normative features cannot vary entirely independently of the world's non-normative features; variation in normative reality requires variation in non-normative reality.

Instead of using terms like "covariation", most philosophers of normativity prefer the term "supervenience". I will follow the terminological tradition and use the term "supervenience" as well. Since I am concerned particularly about the supervenience of normative properties on whatever they supervene upon (or don't supervene upon), I'll also use the term "normative supervenience".

In this section, I only discuss formulations of supervenience that characterize the subvening (base) properties as *non-normative* properties. I will call these formulations "*additive*" formulations of normative supervenience. I call them "additive" formulations because the picture of normativity that emerges from this view of normative supervenience, is as something that is *added on to* non-normative reality–dropped on top of non-normative reality from above, so to speak. On the additive picture of normativity, we can in principle coherently and correctly imagine the following two things: our world, with all its normative and non-normative features; and our world with just its non-normative features.

If it helps, consider the following pair of "normativity glasses". Normativity glasses are just like sunglasses, except for two things. First, unlike sunglasses, wearing them does not filter out sunlight but filters out normativity instead. Second, unlike sunglasses, normativity glasses don't exist. This is a thought experiment.

Normativity glasses work like this. When not wearing the glasses, one's conception of the world is as of possessing normative and non-normative features. But then, when one puts the glasses on, one finds the normative features of the world stripped away, leaving only the non-normative features that characterized one's conception of the world when not wearing the glasses. The additive picture of normativity is a picture of normativity on which the normative sunglasses thought experiment is coherent.

Although I personally find this metaphor helpful, others may find the normativity glasses metaphor to be confusing. Glasses, for instance, can do lots of things to our conception of the world. In addition to stripping away features from the perceptual experience of the world, they can also distort the experience or add additional features. My normativity glasses are doing one specific thing: stripping away from one's conception of the world the world's normative features, leaving the non-normative features intact and undistorted. However, I don't need the normativity glasses metaphor to explain the additive picture of normativity. Enough work is done by the term "additive". Simply put, the additive conception of normativity is a conception of the world's normative features as being features that we can conceive as being added to or subtracted from the objects of the world and their non-normative features.[39]

Additive formulations of supervenience are contrasted with what I will call "transformative" formulations of supervenience, which specify that the subvening properties themselves be normative. Although I cannot talk in more detail about these formulations in this chapter on reasons fundamentalism, I will talk in considerably more detail about them in

[39]Of course, if one accepts that normative reality supervenes on non-normative reality with metaphysical necessity, it would not be possible to add or subtract normative features from the world, leaving the non-normative features of that world intact and undistorted.

§4.[40]

In this section, I'll first talk in a more colloquial fashion about what normative supervenience is. I'll then explain why it is so universally accepted. Then I'll give a more precise characterization of the kind of normative supervenience that is uncontroversially accepted. The purpose of this work is to both motivate normative supervenience as an important philosophical topic and provide a platform for my arguments in §3, where I will argue that reasons fundmentalism cannot explain why the normative features of the world supervene on the world's descriptive features.

Any reasons fundamentalist who accepts supervenience will also be committed to some sort of additive variety of supervenience. Here is Centrality again:

> *Centrality*: For all facts F, F is a normative fact only if F is a reasons fact or F embeds a reasons fact.

A *reasons fact* is just a fact that some other fact is a reason. I mentioned in §2.2 that one potentially surprising implication of Centrality is that most reasons are not normative facts, since most reasons are not facts about some other facts' being a reason. So take again the

> *Tyranny Fact*: Acts of terrorism against citizens of Western countries tend to strengthen tyrannical impulses on the part of Western governments.

We'll suppose again that, given our current global situation, Tyranny Fact is a reason for critics of the West not to commit acts of terrorism against citizens of the West. However, there is nothing about Tyranny Fact itself that has direct implications for how the world

[40]My inspiration for the additive/transformative distinction is an influential pair of papers by Matthew Boyle, who uses a similar distinction in order to contrast two distinct interpretations of how Aristotle conceives of the relation between desires and reason–see (Boyle 2009) and (Boyle 2016).

ought to be or could ought to be. On the reasons fundamentalist picture, a fact has direct implications about how the world ought to be or could ought to be if it is a reasons fact: a higher-order fact that some lower-order fact is a reason. This is what Centrality means. Since Tyranny Fact is not a fact about some other fact's being a reason, it does not itself qualify as a normative fact. It is thoroughly non-normative.

Normative supervenience is a feature of normative properties such that, if something instantiates a normative property, then anything relevantly like that thing also instantiates the same normative property. Therefore, normative supervenience says that, if Tyranny Fact is a reason, then anything relevantly similar to Tyranny Fact will also be a reason. Per hypothesi, Tyranny Fact instantiates the normative property of being a reason. So normative supervenience says that anything relevantly like Tyranny Fact instantiates the normative property of being a reason as well.

Here, finally, is the argument that Centrality commits the reasons fundamentalist to a specifically additive formulation of supervenience. Let the *relevance properties* of Tyranny Fact be the set of properties P of Tyranny Fact such that anything instantiating all the properties in P qualifies as being relevantly similar to Tyranny Fact for the purposes of the supervenience claim. By Centrality, any *normative* property of Tyranny Fact will either be the second-order normative fact that Tyranny Fact is a reason, or some purely formal, higher-order fact built up out of the fact that Tyranny Fact is a reason. Therefore, the only substantive normative property that's available to us to construct the relevance properties of Tyranny Fact is the property of being a reason. If the property of being a reason were part of the relevance properties of Tyranny Fact, then the supervenience claim would be trivial, since it would just state that, if Tyranny Fact is a reason, then anything that has

the various relevance properties of Tyranny Fact and *is also a reason* is a reason as well. Normative supervenience is not a trivial relation. Therefore, the relevance properties of Tyranny Fact need to be all non-normative. If that's true, then normative supervenience says that, if any fact instantiates a normative property, then anything relevantly non-normatively like that thing also instantiates the same normative property. But this just is the general formulation of additive supervenience. Any reasons fundamentalist who accepts normative supervenience at all is therefore committed to some additive variety of supervenience.

What is normative supervenience?

Roughly, A supervenes on B just if any change in A is accompanied by a change in B. You have a lot of H_2O molecules in your Nalgene bottle right now. You also have water in that Nalgene bottle. Now suppose that there were that same Nalgene bottle with those same H_2O molecules in it, but no water were there. Well, that can't be: it's impossible for there to be a Nalgene bottle filled with H_2O molecules that does not also contain water. After all, water just is H_2O molecules.

When we say that it's impossible for there to be a Nalgene bottle filled with H_2O molecules but empty of water, we are making a claim about a certain relation that holds between H_2O molecules and water (in the presence of the Nalgene bottle): that whenever you have H_2O molecules (in the presence of the Nalgene bottle), you will also have water. In other words, the presence of water *supervenes* on the presence of H_2O molecules: you can't have a state of affairs featuring the presence of H_2O molecules without it also being true of that state of affairs that water is present.

In the water/H_2O case, we also have an *explanation* as to why the presence of water supervenes on the presence of H_2O molecules. The explanation is that water *just is* H_2O molecules, where the "is" here is the "is" of identity. Clearly something can't be present unless it itself is present, so it stands to reason that water will be present whenever H_2O molecules are present.

Since we can explain the supervenience present in the water/H_2O case by appealing to the identity of water and H_2O, it also turns out not only that the presence of water supervenes on the presence of H_2O, but that the presence of H_2O supervenes on the presence of water. H_2O and water are the same thing, so whenever you have water present, it stands to reason that you will also have H_2O.[41]

Most cases of supervenience are not symmetrical in this way. In other words, when the presence of something—a fact, property, object; just call it *X*—supervenes on the presence of something else *Y*, it's usually not also true that the presence of *Y* will supervene on the presence of *X*. When an object is blue, for instance, that object must be colored. You can't take away the object's "coloredness" (its property of being colored) without also taking away its blueness. Therefore the coloredness of an object supervenes on the blueness of that object.[42] But it is not true that whenever an object is colored, it must be blue. It could be green or red instead. You can easily take away the blueness of an object (replacing it with redness, for instance) without taking away that object's property of being colored. So

[41] Some formulations of supervenience, such as those found in the philosophy of science, do not allow supervenience to be symmetrical in this way. The metaphysics and metaethics literature, in contrast, tend to include symmetrical varieties of supervenience as genuine kinds of supervenience. I will use supervenience in this latter, more capacious way.

[42] It also supervenes on the object's redness, maroonness and so on, since taking away its coloredness would also require that you take away its greenness, redness and so on.

the object's blueness does not supervene on that object's coloredness.

Normative supervenience is an instance of this more typical, asymmetrical kind of supervenience. When we say that normative reality supervenes on non-normative reality, we mean something like this:

> Let X denote a dictatorial leader. Now think of all the things that X did: sign execution orders for thousands of people, allocate the federal budget overwhelmingly towards military spending, and so on. He was pretty morally bad, no? Now imagine a person Y who is just like X, with just the same psychological history and dispositions as X, placed in the same historical and political circumstances as X, and who made the same choices as X which had the same effects on the world as X's did. That guy Y seems pretty morally bad too, right?

Yes, that guy Y does seem pretty morally bad. If you think that other guy Y is not only morally bad but *has to be* morally bad given the moral badness of X—in other words, if you think it *couldn't be otherwise* that Y is morally bad when X is—you think that moral badness supervenes (on some sense we have not specified) at least given all those other causal, psychological, and historical features described in the quotation above.

Another way to think of normative supervenience is as a constraint on combinations of properties denoted by distinct vocabulary sets. One set of vocabulary–the *normative vocabulary* set–contains terms like "cruelty", "goodness", "honesty", "beauty"; the other set (the "non-normative vocabulary" set) contains terms like "redness", "squareness", "heat", "immobility". The supervenience constraint is simply that any two objects that share the same set of non-normative truths also share the same normative truths. To put it another way, suppose that there are two numerically distinct objects x and y such that any non-normative truth about x is also a non-normative truth about y, and any non-normative truth about y is also a non-normative truth about x. Normative supervenience would then entail that any

normative truth of *x* is a normative truth of *y*, and any normative truth of *y* is a normative truth of *x*.[43]

This, very roughly, is normative supervenience. It is, however, another question entirely why people do and should accept it. In the next subsection I'll discuss why people accept it, and I'll follow that subsection with a precise formulation of what it is that metaethicists seem to be minimally accepting when they accept normative supervenience.

The least controversial doctrine

There can be no normative difference without a non-normative difference–at least, that is what we say. Normative supervenience is a doctrine that's affirmed by everyone in metaethics: realist irreduciblists[44] as well as realist reduciblists,[45], particularists[46] as well as generalists, and even the wide array of normative non-realists: error theorists (Olson 2012, p. 92), subjectivists (Smith 1994, pp. 21-22), constructivists (Meyers 2012, p. 12) and expressivists[47].

So uncontroversial is the doctrine of normative supervenience that its being uncontroversial is nearly as uncontroversial as the view itself.[48] Thus Michael Smith claims, "Every-

[43] I myself think that this way to understand supervenience is somewhat misleading, since it construes the supervenience constraint as a semantic constraint rather than a metaphysical one, a constraint on admissible descriptions rather than a constraint on admissible property instantiations. I've found, however, that this way of explaining supervenience has helped some interlocutors understand the phenomenon, so I include it here.

[44] e.g. (Shafer-Landau 2003, p. 77), (Enoch 2011, pp. 136-137), (Scanlon 2014, p. 33), (Parfit 2011, p. 756), (Kaspar 2012, section on "Supervenience Problems").

[45] e.g., (Jackson 1998, p. 118)

[46] (Dancy 2004, pp. 86-7), (Strandberg 2008, pp. 129-158), (McDowell 1998, p. 202), (McNaughton 1988, p. 62).

[47] (Hare 1984), (Blackburn 1993), (Gibbard 2003).

[48] (Shafer-Landau 2003, p. 77), (Majors 2009, p. 29), (Sturgeon 2009, p. 53).

one agrees that the normative features of things supervene on their natural features" (Smith 1994, p. 21) and that "everyone agrees that this is a platitude, an a priori truth" (ibid., p. 22). Michael Ridge calls the supervenience of the ethical on the descriptive "a basic platitude" (Ridge 2007, §6) and that "someone who denied a suitably formulated supervenience thesis would thereby give evidence that he is not a fully competent user of normative terms" (ibid., p. 331).[49]

I have long found these claims disturbing. Philosophers are contentious folk, particularly in the domain of moral theory; controversy over doctrines tends to be the norm rather than the exception. Uncontroversy is unusual—even, we might say, a queer feature of our metaethical discourse, one in need of explanation and defense. And yet, to my knowledge, there are only two philosophers who explicitly state in print that they do not accept additive supervenience. The first is Gideon Rosen, who is moved by considerations of neo-essentialism to call supervenience into question. The second is Akeel Bilgrami, who argues that there exists no coherently formulable statement of a normative supervenience doctrine. I will discuss both of these views extensively in the later chapters of this dissertation.

Why does nearly everyone accept supervenience? Here are some reasons.

Supervenience and intuitions

The dictator illustration that I gave in the last section aimed to pump your intuitions about normative supervenience. If you thought the dictator duplicate was morally bad on the

[49]Nicholas Sturgeon argues that normative supervenience is only uncontroversial as a general doctrine, but that, when one teases apart the specific formulations that various theorists commit themselves to, we find that there is no single precise formulation that everyone agrees with. However, the various formulations Sturgeon describes concern how exactly to specify the nature of the non-normative reality that normative reality is supposed to supervene on. He does not dispute that nearly everyone accepts that normative reality supervenes on non-normative reality in some general sense.

assumption that the real-world dictator was, then you have the normative supervenience intuition. And if you have this intuition, you are also part of the very large majority of philosophers who also say they share the intuition, because the vast majority of philosophers say they do.

Non-philosophical folk may have the intuition as well. Frank Jackson, a renowned champion of the folk, remarks that "the most salient and least controversial part of folk moral theory is that normative properties supervene on descriptive [non-normative] properties, that the ethical way things are supervenes on the descriptive]non-normative] way things are" (Jackson 1998, p. 118). Since Jackson makes this claim without any additional elaboration, it is not clear what Jackson means by this or why he thinks it is true. He has not, as far as I know, measured reactions from non-philosophers by means of survey questionnaires or magnetic resonance imaging. My best guess is that he means that the folk as well as the professional philosophers have the intuitions that underlie the dictator case above. This, I take it, is probably true.

My take is that intuition-based defenses of philosophical positions are fine as they stand, but they at best provide only defeasible support in favor of philosophical positions. Intuitions are subject to the tribunal of philosophical reflection. For instance, we can consider a hypothetical error theorist, who might say, "Sure, Hitler seems pretty bad, but it turns out that he's not, because after all we have good theoretical reasons to think that there are no instantiated normative properties at all." Intuitions also face the tribunal of psychologically-based etiological debunking, as Joshua Greene's work as been keen to show (Greene 2007).

The second argument in favor of normative supervenience—the argument from folk morality—is surprising. Supervenience, first of all, is a fairly obscure topic, and one that

did not even merit much philosophical discussion until the past several decades. It is unlikely that, as Brink says about moral motivational internalism, "any belief so recherche could be part of common sense moral thinking" (Brink 1984). Second, folk morality is notoriously a mess of conflicting intuitions (Sverdlik 1985). In their attempts to impose some systematicity into folk morality, there seems to be no aspect of it that moral philosophers have left unchallenged. The utilitarian's struggles with folk theories of love and friendship, the error theorist's dismissal of ethical truth—these are just two examples of the lengths to which ethical theorists have been willing to pursue their theory at the expense of folk morality. Even Jackson allows for some moral revisionism, since his own theory appeals to the notion of a mature folk morality that can be achieved only through a process of moral revisionism. One must wonder why supervenience has largely been immune to this sort of revisionism, and that supervenience would survive the modifications that would need to be made to actual folk morality for it to blossom into maturity.[50]

Normative explanation and evidence

In my view, the stronger defense of normative supervenience appeals to features of normative explanation. Let's return again to the dictator case (last time, I promise). Suppose hypothetical dictator *Y* were not morally bad. Perhaps *Y*'s actions were not so horrific, or he lacked political power, or maybe he was operating under some fundamental misconceptions about the laws of the universe–perhaps he thought that sentencing people to concentration camps would bring them pleasure or save their souls or what have you. If there were one

[50]Although there have been a number of objections raised against Jackson's argument in chapter 5 of (Jackson 1998), no one to my mind has raised the objection that the maturation process to which Jackson appeals could result in the rejection of normative supervenience.

of these differences, then one might be able to make a case that although dictator X were morally bad, Y would not be. However, we have stipulated away that there are any such differences. If X were morally bad and Y were not, this would be the only difference between X and Y. X is morally bad because X *just is*; Y is not morally bad because Y *just is not*.

The example I gave was in terms of moral badness, but similar examples can be given for reasons for action. Here again is the

> *Tyranny Fact*: Acts of terrorism against citizens of Western countries tend to strengthen tyrannical impulses on the part of Western governments.

Tyranny Fact is a reason for critics of the West to avoid engaging in acts of terrorism. Now imagine a world just like ours, with just the same history and same political leaders with the same psychological dispositions as our world, with the same causal laws as those that obtain in our world. The only difference is that in that world, Tyranny Fact is not a reason for critics of the West to avoid engaging in acts of terrorism. Tyranny Fact *just is* a reason in our world, and *just is not* a reason in that other world we just described, and that's all there is to it. This would be really odd, and probably incomprehensible.

One way to understand the importance of supervenience is to contrast the normative case to the case of phenomenal experience. In inverted spectrum thought experiments, we are asked to imagine two people who are exactly alike, down to the last neurological molecule, but different in the following way: when one of the people perceives some object, she has a red experience; but when the other person perceives the same object in exactly the same viewing conditions, she has a green experience.

This kind of phenomenon could turn out to occur in our world, or not; but whether there exist inverted spectra is not the point of my use of this thought experiment. What I want to call attention to is the fact that an inverted spectrum situation is a conceivable situation. We can imagine what it would be like for us, say, to remember our experiences and the red and green content of these experiences; and then imagine that, in all these experiences, the originally red content of the experiences were instead, green, and the originally green content of the experiences were instead, red. There is nothing about the nature of red and green that prohibits us from conceiving of our experiences in this way.

Normative properties may also admit of this same inversion. I do think it's possible for us to get into a mindset in which a dictator that signed execution orders for thousands of people and overwhelmingly allocated the federal budget towards military spending was morally good. After all, plenty of people have thought that particular dictatorial leaders that have done these kinds of things in the past were morally good for doing so. Placing ourselves in this mindset, however, reveals an important disanalogy between the phenomenal and normative cases. It is conceivable and comprehensible to say that two physically indiscernible people, with the same history, in the same viewing conditions, could have experiences of the same object as being of a different color. Metaphysically speaking, the color experiences of people can, in principle, vary independently of the physical nature of the experiencer, object, and viewing conditions. But it is neither conceivable nor comprehensible to say that two physically indiscernible people, with the same history, in the same situations, could differ in their moral qualities. The reason for this is because normative properties, unlike phenomenal properties, are conceptually such that, for normative properties to be instantiated in an object or person at all, there has to be an explanation of their

instantiation, there has to be something else about that object or person that can serve to explain why that object or person has the normative properties that she does. In the case of color properties, there may very well be an explanation as to why certain experiences have these properties and certain other experiences do not, but this is not a conceptual constraint given by the nature of color properties.

To put the point another way, imagine again two physically indiscernible people with the same history, in the same situation, looking at the same thing. You know that the first person is experiencing the object as being red, and we also know that this same first person is morally good. What color experience is the second person having? Probably a red experience. But you could be wrong. It could be that whatever laws that govern the relation between physical and phenomenal properties are statistical or otherwise "chance-y"; it could be that no laws govern the relation between physical and phenomenal properties at all. There is nothing about phenomenal properties themselves that prohibit this kind of variation of phenomenal properties despite a lack of physical difference or situational difference. Normative properties do not work this way. It makes no sense to say that the first person is morally good but the second person is not, despite there being no other difference between the two people and their situations that could explain the difference.

It's true that some states of affairs have no explanation. They just are the way they are. One example is the fundamental gravitational constant. Most of us (i.e., everyone but necessitarians) think that it is metaphysically possible that the gravitational constant could have been $6.675x10^{-11}Nm^2/kg^2$ rather than $6.674x10^{-11}Nm^2/kg^2$. Perhaps, as some people try to do, the value of this fundamental constant can be explained by appealing to

God or the multiverse hypothesis.[51] Or perhaps, as Parfit says and I am inclined on most days to believe, "It is random that reality is as it is... it has no explanation of any kind" (Parfit 1998).[52]

Another example is the particular collapse of the wave function from superposed to classical states. Why did the photon "land" in that place rather than one of the many other places when we shot it at the double slit? Well, the wave function just collapsed that way. Nothing more to be said. We can offer a partial explanation as to why the wave function collapsed the way it did. We can say, for instance, why it landed on the photographic film and didn't quantum tunnel to the Andromeda galaxy instead: the first is much more probable. But why it "landed" where it did rather than a little further away? Nothing more to say.[53]

Although there are these descriptive features of our world that are at least to some extent explanatarily opaque, and this is somewhat puzzling and concerning, it would be much more puzzling and concerning if the moral features of our world were similarly devoid of explanation. I'll give two reasons to think this: one epistemological, the other metaphysical

[51] The appeal to God, to me, seems problematic just because there would still be a question as to why God, who *per hypothesi* wants to bring about rational life, would pick this gravitational constant rather than the actual gravitational constant plus ϵ, where ϵ just is a small enough number that the difference would not affect outcomes in terms of the development of rational life. So there would still remain the question "Why this constant rather than some other"? There is a kind of multiverse hypothesis that could explain why our universe has the gravitational constant that it does because it can state that for every positive real value x, there is a universe featuring fundamental laws that are exactly like ours except that they have x as their gravitational constant.

[52] There is also the view, shared by Russell, van Fraassen and others, that the question whether the universe has an explanation is incoherent. I think it's perfectly coherent, but I'm discussing explanations of the universe here purely for illustrative purposes and nothing argumentative in the dissertation hangs on it. To the extent that it is relevant, the incoherence of the question might give more weight to the claim that there are no brute (explanationless) facts, which actually helps my argument in §3 that the reasons fundamentalist has to claim that the fact of additive supervenience is a brute fact.

[53] A minority of theories of quantum phenomena are fully deterministic theories and therefore would deny that there is "nothing more to say".

and conceptual.

First, in the empirical cases of brute facts or brute differences, we can use our senses to tell whether the fundamental constant is one way or the other and whether the wave function collapsed in one way rather than the other. The object would strike the ground at a different time; the photographic film would have been marked in a different place. In the case of the wave function, there are very many potential collapses of the wave function for any fully determinate prior setup of the double slit experiment, so the collapse does not supervene on these causal antecedents. In the case of the fundamental constants, it is strange to even think about what the "prior setup" would even be, so the supervenience of the values of the fundamental constants on the "prior setup" will probably turn out to be incoherent. The point is that, because there is a failure of supervenience, there are no laws that we can apply to the fully determinate prior setup and infer a priori what the fundamental constants will be or how the wave function will collapse. We have to rely on further empirical evidence.

No such empirical tests exist to detect brute normative difference. If we did have a quasi-perceptual faculty of moral sense, we could appeal to this faculty to justify our belief that *X* is morally bad but *Y* is not. However, almost no one thinks we have a quasi-perceptual moral sense.[54] Moore is the philosopher most characteristically derided for believing in the quasi-perceptual moral sense faculty, but even Moore would deny that you could use this faculty to tell which of the two non-normatively indiscernible tyrannical dictators was morally bad and which was not. This is because Moore believed in a variety of normative supervenience. He writes:

[54]The only major published exception to this that I know of is (Oddie 2009).

> If a thing is good (in my sense), then that it is so *follows* from the fact that it possesses certain natural intrinsic properties, which are such that from the fact that it is good it does not follow conversely that it has those properties (Moore 1922, p. 261).

Without a quasi-perceptual moral sense with even more powerful capacities of epistemic discrimination than the intuitive moral sense faculty that Moore thought we had, there is nothing we can appeal to in order to justify any belief that failed to respect a supervenience constraint. For instance, if we believed that tyrannical dictator *X* were morally bad while his indiscernible duplicate were not, and we had no intuitive or quasi-sensory capacity to justify this belief, it's not clear to me what other epistemic capacity of ours could provide justification for this belief. Normative supervenience then becomes a constraint on warranted belief-formation about normative property instantiations for those of us who do not want to accept that we have this very robust, quasi-perceptual faculty of moral sense.

Secondly, I want to argue that we need some sort of normative supervenience doctrine in order to have normative *explanation*. Since we have reason to accept that there is such a thing as normative explanation, we therefore need normative supervenience.

Take two utilitarians[55] who are arguing about how best to allocate their resources. As utilitarians, they agree that what you ought to do is engage in activities that alleviate the most pain. One of them has recently been considerably moved by a speech given by Peter Singer and is arguing that their resources morally ought to be donated to OxFam. The other responds by citing her family situation. She indicates that her family is poor and needs financial support; she thinks it would be morally better for her to transfer financial assets to them. They both agree on the relevant non-normative facts: that donating resources to

[55] Specifically, non-scalar objective act utilitarians.

OxFam will alleviate dehydration and that donating resources to the second utilitarian's family will alleviate their poverty. Now, the first utilitarian might say to the second, "I agree that the situation of your family is quite bad. But consider OxFam's recent campaign in Afghanistan, where they are distributing water to people in areas who lost access to it because of war and drought. Your family is poor, but they at least own a run-down house and have limited governmental support. Starvation and dehydration are much more painful than that. You are in a difficult situation because, after all, they are your family. But since you are a utilitarian, it's clear what you morally ought to do."

The first utilitarian in this case is giving an explanation as to why donating to OxFam is the morally right thing to do. We can think of it as a kind of deductive explanation, proceed from a major normative premise (You morally ought to do things that alleviate the most pain) and minor non-normative ones (dehydration and starvation is the most painful thing; donating to OxFam alleviates starvation and dehydration), to a conclusion about what you morally ought to do (donate to OxFam). The point of bringing this up is not the details of this argument, which most of us I suspect would find either morally objectionable or empirically misinformed. The point is rather to illustrate a feature of the major premise in the argument. If the major premise holds, then it follows that, for any set consisting of the possible courses of action that are open to you (call these *options*)[56] and their consequences, the facts about which options are morally right will be "fixed" by the facts about the various options' consequences. Call these sets of {option, consequence} pairs *practical situations*. Given the major premise, any two non-normatively indiscernible practical situations will

[56]The notion of a practical option for an agent is much more involved than is indicated by my use of it here. For more discussion of this, see (Rovane 2013, pp. 208-9) and (Levi 1986).

also be indiscernible with respect to the facts about which options, given the practical situation, are morally right. But this is a supervenience claim. Moral explanations that are given on the basis of major premises that bridge non-normative and moral facts therefore presuppose a supervenience relationship between the non-normative facts and the moral ones.

Here are some more remarks that may help elucidate the connection. We can think of the major premise serves as a function from practical situations to facts about moral rightness: tell it which option alleviates the most pain, and it will tell you which option (or set of options, in the case of options whose consequences alleviate identical quantities of pain) is morally right. Even if it's indeterminate which option alleviates the most pain, the function would give us the result that it is indeterminate which option is morally right. In order to get a different result, you would need to change at least the facts about it being indeterminate which option alleviates the most pain. In other words, the output of the function supervenes upon its input.

The above remarks aimed to show that we do have a practice of providing normative explanations. Furthermore, this is not a practice that we can easily give up.[57] Here is a very helpful long quotation from Nick Zangwill that expresses the importance of normative explanation fairly clearly:

> Suppose... that I say that Billy is bad. You ask why. I reply, "No reason; he's just bad, that's all." This is utterly irresponsible and weird... For if Billy is bad, there must be something that makes Billy bad. We must think that Billy is bad because or in virtue of the way he is in other respects... Mackie was right to mention, if only in passing, the problem of this because ("Just what in the

[57]Even error theorists are hesitant to give up practices of normative explanation. For more on this, see the ending remarks on moral conservatism in (Olson 2012).

> world is signified by this 'because'?" (Mackie 1977, p. 41)). Different theories are possible here.... Nevertheless, everyone should agree on the existence and centrality of this "because". Moreover, this because is not just metaphysical constraint on properties but also a constraint on our judgments (a "conceptual" feature of them). The slogan might be: not just bad, but bad because: we judge not that something is bad period, but that it is bad because of certain natural properties. It is a priori that moral properties depend. Let us call the requirement to judge that something is M [moral] because it is N [non-normative] the 'Because Constraint'. (Zangwill 2006, pp. 270-271)

There are two components to Zangwill's claim. The Because Constraint is, first, a metaphysical constraint: normative properties cannot be instantiated in an object or state of affairs etc unless that object or state of affairs possesses some other, non-normative property in virtue of which it also possesses the normative property. Secondly, it is a conceptual constraint: what it means for a property to be a normative property is partly for that property to be instantiated in virtue of other, non-normative properties. Since the "Because Constraint" is constitutive of (entailed by) any normative concept, the Because Constraint is therefore a constitutive part of our practices of normative judgment. It makes no sense, for Zangwill, to call a judgment a normative judgment unless it is a judgment that something has the particular normative property you're judging it to have in virtue of some other, non-normative property.[58]

Commitment to the Because Constraint entails commitment to some sort of normative supervenience doctrine. Suppose Billy is bad in virtue of his disposition to bully other kids on the playground. That's to say that Billy's disposition to bully fully explains Billy's bad-

[58] Zangwill recognizes that sometimes we make normative judgments on the basis of testimonial evidence. So if the Dalai Lama says that being greedy is wrong, and we believe that being greedy is wrong on this basis, this testimonial evidence is not a feature of the greediness that we have judged is wrong. However, Zangwill thinks that we do *indirectly* satisfy the Because Constraint even in this case, because it's part of our deference to the Dalai Lama that we take ourselves to have good reason to think that the Dalai Lama is making the moral judgment about greediness on the basis of some non-normative feature that greediness possesses.

ness given the nature of badness. Now, Billy's friend Charlie has the same disposition to bully kids on the playground. Charlie's bullying would then provide us with a full explanation of Charlie's being bad. And if Charlie's bullying also gives us a full explanation of Charlie's being bad, then Charlie is bad. Ditto for any other child who has the same disposition to bully. Badness then would supervene on childhood dispositions to bully, since you can't have a child who has the bullying disposition without that child being bad as well.

To see this, here is an argument from contradiction. Suppose Charlie had Billy's same disposition to bully, but Charlie was not bad. Charlie grew up in an abusive household and developed depression and ADHD at an early age. Charlie is not bad; rather, he is a victim of mental illness brought about by domestic abuse. Billy, in contrast, had the best upbringing imaginable and just chooses not to be very nice. What this reveals, I think, is that our previous explanation of Billy's badness in terms of his dispositions to bully was not a full explanation. It left something out. Billy is not bad simply because be is disposed to bully kids on the playground. He is bad because he has this disposition *and* the disposition is not the result of mental illness. Once our previously incomplete explanation of Billy's badness is expanded into a full explanation, Charlie turns out not to have properties that provide a full explanation for the presence of badness in Charlie after all.

The existence of a full explanation for the presence of a normative property therefore entails that the presence of that normative property supervenes on the presence of the features that constitute the full explanation. We do of course offer explanations like the following: the accident happened because the driver was drunk and there was ice on the road. Many drivers drive drunk on icy roads and don't cause accidents. So the presence of an accident does not supervene on the presence of an intoxicated person driving on an icy road. This is

not, however, a counterexample to the claim that full explanations entail supervenience. It only reveals that most of our explanations are not full explanations. In this icy road explanation of traffic accidents, lots of salient details were left out: the presence of other drivers on the road, the absence of guardrails, the precise nature of the intoxication and distribution of ice and so on. It would take a huge amount of effort to come up with the full range of features that explain road accidents. We don't usually invest this effort because the benefit of coming up with a full explanation of road accidents does not outweigh the huge costs. In the road accident case, incomplete explanations are good enough. We know that intoxication and icy road conditions are strong predictors of road accidents. This knowledge provides us with enough of a justification for public policies such as funding for ice removal and laws against drunk driving.

Normative arbitrariness

Lest you are not already convinced that we should accept some sort of normative supervenience doctrine, here are a few closing remarks. The first remark is from Shafer-Landau, who writes, "If the moral fails to supervene on the non-moral, then the non-moral world does not control the moral world. But if that world does not control the moral world, then the moral world is out of control. Moral assessments would be arbitrary" (Shafer-Landau 2003, p. 77). Normative supervenience (of which moral supervenience is a subset) allegedly anchors the moral "world" to the non-normative world so that the moral world is not tossed about by the tides of moral contingency.

And moral contingency is very bad. Consider Warnock, who writes, "The picture presented [in Warnock's presentation of Moorean intuitionism] is that of a realm of moral

qualities, sui generis and indefinable, floating, as it were, quite free from anything else whatever, but cropping up here and there, quite contingently and for no reason, in bare conjunction with more ordinary features of the everyday world" (Warnock 1967, p. 14). Although one (for instance, myself) might draw issue with the description Warnock gives here of Moore's views, the passage describes a characteristic that that I take it moral philosophers across the board would want their preferred theory to lack. We don't want the moral features of the world to "float quite free from anything else whatsoever." If normative supervenience is false, we would, for instance, have to take seriously the thought that, even though act utilitarianism is false and that it is therefore not morally prohibited for me to be working on this paper right now instead of wiring all the many tens of dollars in my possession to the Bernie Sanders campaign, quitting my program and joining an international volunteer relief organization, the conjunction of the natural features of this world and the truth of act utilitarianism may in fact be possible. It could be that act utilitarianism is true in a non-normatively indiscernible world occupied by my trans-world counterpart.

This thought, from conversations that I have had, fills many people with apprehension. The sources of this apprehension are both epistemological and personal. In terms of moral theory, the denial of metaphysical necessity raises the question how we are to know that we aren't after all in the act utilitarian world. We can't appeal only to descriptive features of the world to explain why moral reality is how it is rather than some other way, since our world and the act utilitarian world are stipulated to be descriptively indiscernible.[59]

The epistemological worry also raises a personal concern. Most of us who do not work on Wall Street want to be morally good people; we want, according to the moral realist,

[59] A more rigorous investigation of these issues is provided in (Street 2008).

to instantiate the property "morally good". If act utilitarianism were true, I for one would not qualify as a morally good person. I may be contributing to running classes that raise social awareness and concern, either directly or indirectly, among a few hundred students. I may in the future write popular books that produce beneficial social effects across an even wider audience. Even in this ideal case, however, the fact remains that there is a surfeit of very intelligent and highly skilled philosophers who would happily take any such job and do just as well at it, and a paucity of very intelligent people willing to work with the poorest and most exploited populations whose needs (including the need to think about philosophical issues!) are so immediate and penetrating.[60] If there turns out not to be an adequate solution to the epistemological worry, we can continue our lives in the hope that act utilitarianism is false. But we may be wrong. Even if we are correct, our belief in the falsity of act utilitarianism may fall short of the certainty characteristic of knowledge, and the apprehension that would result for those of us that care about being good and doing the right thing would be something that most of us would want to avoid.

Clarifying normative supervenience

Supervenience relations are well-known and have been discussed and distinguished from each other at endless length. In this section, I'll be using that work to argue for the following more precise characterization of normative supervenience which I take most metaethicists to implicitly or explicitly endorse:

[60]These thoughts are somewhat personal in nature, but I think that thoughts like this are highly appropriate in the context of a dissertation in moral philosophy: an area of philosophy whose subject matter is partly constituted by issues of broad personal concern. I also imagine, too, that many moral philosophers have at one time or another entertained thoughts like this—a fact that would, if true, underscore the relevance of these remarks.

> *Full Strong Individual Additive Supervenience*: For any possible worlds w_1 and w_2 and any individuals i_1 in w_1 and i_2 in w_2, if i_1 and i_2 are intrinsically and extrinsically non-normatively indiscernible and w_1 and w_2 are non-normatively indiscernible, then, for any normative property P, i_1 instantiates P if and only if i_2 instantiates P.

Since the reasons fundamentalist thinks that all that is normative consists in instantiations of the favoring relation and purely formal higher-order features of instantiations of this relation, the relevant application of Full Strong Individual Additive Supervenience for the formulation of reasons fundamentalism is

> *Additive Reasons Supervenience*: For any possible worlds w_1 and w_2 and any fact-agent-action-circumstance tuples T_1 in w_1 and T_2 in w_2, if T_1 and T_2 are intrinsically and extrinsically non-normatively indiscernible and w_1 and w_2 are non-normatively indiscernible, then T_1 instantiates the favoring relation if and only if T_2 instantiates the favoring relation.

I'll then be in a position to exploit Additive Reasons Supervenience in the argument presented in §3 and §4.

I'll start with a distinction between weak and strong supervenience that we owe to Jaegwon Kim (Kim 1987).[61]. If *B* weakly supervenes on *A*, then you can't have a *B* difference without an *A* difference. This is also true for strong supervenience: if *B* strongly supervenes on *A*, then you can't have a *B* difference without an *A* difference. The difference between weak and strong supervenience has to do with the way we quantify over possible worlds in the supervenience formulation—or, alternatively, with the way we introduce modal operators. Weak supervenience is supervenience within a world, while strong supervenience is supervenience across worlds. If *B* weakly supervenes on *A*, then, for any world *w*, anything

[61] See also (Davidson 2001, pp. 170-183)

that's A-alike in w will also be B-alike in w. If B strongly supervenes on A, then any two things that are A-alike will also be B-alike, no matter what world they're in.

Here are some examples. In relativity theory, the energy of a universe weakly supervenes on the total mass of that universe. This is because the total energy of a universe is a function of the mass of that universe and the speed of light. However, the energy of a universe does not strongly supervene on the total mass of that universe, because there are metaphysically possible universes in which relativity theory holds but the speed of light is different. Tallness also weakly supervenes on height. Any two people in a given world who are 6'5" in height will also be alike with respect to whether they are tall. In this world, they are. But tallness does not strongly supervene on height, since people who are 6'5" in height will not be tall in worlds where Yao Ming's height is just average.

Now for some examples of strong supervenience. The amount of heat in a system strongly supervenes on the amount of mean kinetic molecular energy of that system since heat necessarily is mean kinetic molecular energy. Whenever you have a newton of force acting on a kilogram of mass, you will have one joule of heat, since that is what a joule of heat is. Although the tallness of any person, as previously mentioned, does not strongly supervene on that person's height, the distribution of tallness across a population strongly supervenes on the distribution of heights across that population. In any world with the same distribution of heights as ours, Yao Ming will be tall; in any world in which the median height of the population is 10 feet but Yao Ming is the same height as he is in our world, Yao Ming will not be tall.

In talking about supervenience so far, I've helped myself to language about possible worlds. This is because supervenience relations are modal relations, and a natural way

to talk about modality is to talk about it in terms of possible worlds. Most discussions of supervenience in metaethics help themselves to possible worlds talk, and I will comply with the conventions of my specialization's discourse in adopting this way of speaking as well. In saying this, I am *not* committing myself to the following Leibnizian conditionals in any robust sense:

1. For all facts F, F is necessary $\iff$ F exists in all possible worlds; and
2. For all facts F, F is possible $\iff$ F exists in a possible world.

The reason I'm not committing myself to these conditionals is that I am neutral for the purposes of this dissertation on the question whether there are possible (non-actual) worlds, whether abstract or concrete, and therefore neutral on the question whether facts exist in these non-actual possible worlds. I am happy to agree, if the reader prefers, that talk of possible worlds is not more than a useful way of talking about modality (Rosen 1990).[62]

Here is what weak supervenience looks like when applied additively to the case of normative supervenience:

> *Weak Individual Additive Supervenience*: For any possible world w and any individuals i_1 and i_2 in w, if i_1 and i_2 are intrinsically non-normatively indiscernible, then, for any normative property P, $P(i_1)$ if and only if $P(i_2)$.

Weak Individual Additive Supervenience says that, if you take any two individuals in the same world and those two individuals are exactly alike in all their intrinsic non-normative properties (i.e. they are non-normative duplicates), then they will also be alike in all their normative properties.

[62] I do think, though, that possible worlds talk is a useful way of talking about modality. I'm not neutral on *that*.

Weak Individual Additive Supervenience is called *Weak* Individual Additive Supervenience because it is strictly weaker than (it is entailed by, but does not entail) the following relationship:

> *Strong Individual Additive Supervenience*: For any possible worlds w_1 and w_2 and any individuals i_1 in w_1 and i_2 w_2, if i_1 and i_2 are intrinsically non-normatively indiscernible, then, for any normative property P, $P(i_1)$ if and only if $P(i_2)$

Strong Individual Additive Supervenience says that, if you take any two individuals in two worlds and those two individuals are exactly alike in all their intrinsic, non-normative properties (i.e. they are non-normative duplicates), then they will also be alike in all their normative properties, no matter whether they are in the same world or a different world. Strong Individual Additive Supervenience entails Weak Individual Additive Supervenience because Weak Individual Additive Supervenience is what you get in those instances of Strong Individual Additive Supervenience in which $w_1 \approx w_2$.

Both Weak Additive Individual Normative Supervenience and Strong Individual Additive Supervenience specify that the subvening properties are the *non-normative* properties. As mentioned in the introduction to this section on supervenience, I will call all such formulations "additive" formulations of normative supervenience. I call them "additive" formulations because the picture of normativity that emerges is as something that is kind of dropped on top of non-normative reality from above.

Both varieties of supervenience are formulated in terms of quantification over possible worlds. There are also formulations in second-order logic extended with modal operators. Here they are:

1. $\Box\forall x\forall F \in A[Fx \rightarrow \exists G \in B(Gx \wedge \forall y(Gy \rightarrow Fy))]$

2. $\Box\forall x\forall F \in A[Fx \rightarrow \exists G \in B(Gx \wedge \Box\forall y(Gy \rightarrow Fy))]$

Formulation (1) is a formulation of weak supervenience. It says that if property set *A* weakly supervenes on property set *B*, then, necessarily, if anything *x* has some property *F* in *A*, then there is at least one property *G* in *B* such that *x* has *G*, and everything that has *G* has *F*. Formulation (2) is a formulation of strong supervenience. It says that if property set *A* strongly supervenes on property set *B*, then, necessarily, if anything *x* has some property *F* in *A*, then there is at least one property *G* in *B* such that *x* has *G*, and *necessarily* everything that has *G* has *F*. The only difference between the two formulations is the addition of the second necessity operator.

There are some issues regarding equivalency relations between the quantification and modal formulations. In particular, (Kim 1987, pp. 79-81) had maintained that the quantifier formulations were equivalent to the modal formulations; this was disputed in (McLaughlin 1995). I am setting these issues to the side and using the quantification over possible worlds formulations because I think these formulations will be clearer to readers.

Neither the weak nor the strong formulations of individual normative supervenience that I have just given are ecumenical. Hare explicitly rejects Strong Individual Normative Supervenience (Hare 1984, p. 4) in favor of the weak formulation. Hare argues for what he calls *universal prescriptivism*: the view that moral judgments such as "it's wrong to take tips out of the cashier's tip jar" are preferences that express principles that one takes to be overriding and universal (Hare 1982, pp. 20-24, 107-16). By "overriding", Hare means that, when one expresses a moral judgment, one's preference to act in accord with

the principle outweighs whatever other preferences one might have. By "universal", Hare means that moral judgments express principles that one prefers everyone to obey, including oneself. If I judge that it's wrong to take tips out of the cashier's tip jar, I prefer not only that everyone else not take these tips but that I not take these tips as well. Hare's universal prescriptivism entails that wrongness weakly supervenes on those properties of acts that one overridingly and universally prefers never to be done. For any two acts, if they are both tip-stealing acts, then they are both wrong. However, wrongness does not strongly supervene on the property of being an act of tip-stealing, since one's counterparts can have different preferences in different worlds, in which case, in those worlds, tip-stealing would not come out as wrong.

Particularists deny both weak supervenience and strong supervenience in favor of global supervenience,[63] which we can formulate for the normative case as follows:

> *Global Additive Supervenience*: For any two worlds w_1 and w_2, if w_1 and w_2 are non-normatively indiscernible, then w_1 and w_2 are normatively indiscernible.

Particularists deny weak and strong individual supervenience because they hold that the moral features of an object can always, in principle, be altered by the addition of new situational features that may, in Dancy's terminology, enable, disable, intensify or attenuate the moral properties in question. Enablers and disablers are features of a situation the introduction of which makes a fact a reason, or makes a reason just a mere fact after all. An intensifier or attenuator is a feature of a situation that strengthens or weakens the strength of a reason, or the weight that we ought to assign the reason in our reasoning practices. The

[63] Dancy used to deny global supervenience, but he now accepts it at (Dancy 2004, p. 87). I can't think of other exceptions to the general claim that particularists accept global supervenience.

goodness of giving to the Red Cross missions in Uganda can be intensified by the existence of a wealthy donor who has pledged to match some Red Cross donations; it could be attenuated or disabled if the leader of Uganda has plans to confiscate some of the aid sent to Red Cross in that country, and so on.

One can easily avoid the particularist worries mentioned above simply by going with a global formulation of supervenience.[64] However, global formulations are defective in that they do not do the best justice to some of the reasons we want normative supervenience in the first place. The intuition pumps in §2.5.1 and §2.5.2 were all of the following form: imagine two individuals, indiscernible in all their intrinsic features and all relevant extrinsic ones. The first individual has such-and-such normative property. Does the second one have that same normative property as well? A "yes" answer to this offers support for formulations of supervenience that place these individuals at the center of the formulation. Global formulations of supervenience don't do this; only individual formulations do. Furthermore, the "Because Constraint" is the constraint that individuals instantiate normative properties in virtue of their non-normative properties. Individual formulations of supervenience are again more suited to capture this idea. In other words, we don't want a formulation of supervenience in terms of distribution of normative properties across non-normatively indiscernible worlds, but a formulation of supervenience, suitably constrained, in terms of distribution of the normative properties of non-normatively indiscernible individuals. So instead of jumping ship on the individual formulations in order to accommodate particular-

[64] I want to accommodate the particularist because particularism is consistent with reasons fundamentalism. Reasons fundamentalism is a theory about the nature of reasons, while particularism is a theory of normative explanation. I am looking for a formulation of normative supervenience that characterizes reasons fundamentalism in general, not just reasons fundamentalisms that reject the normative explanatory views of the particularist.

ists, we should instead see if we can produce an individual formulation of supervenience that answers to the particularist worries.

We can respond to the particularists' worries by expanding the requirements for individuals to qualify as "intrinsically non-normatively indiscernible" to include the requirement that they be indiscernible with respect to all the potential enablers, disablers, attenuators and intensifiers. Since, for the particularist, there is in principle no non-trivial limit to the number of facts (property instantiations) that could play this role, our ecumenical formulation of normative supervenience should take all possible such properties into consideration.[65]. In the Red Cross case, we should understand non-normative indiscernibility in such a way that two acts of donating to the Red Cross mission in Uganda are not non-normatively indiscernible in the relevant sense unless these acts are also indiscernible with respect to relevant extrinsic properties, such as their properties of being such that the leader of Uganda is not corrupt. Really, in order to accommodate the most radical particularist, we should restrict our supervenience claim to only those individuals that are indiscernible with respect not only to their intrinsic non-normative properties, but also indiscernible in the sense that they be identically situated in non-normatively indiscernible worlds. We want, in other words, to build into our criteria for indiscernibility the requirement that there not be any more space in a world for a non-normative property to turn up and enable/disable/attenuate/intensify the normative properties that an individual has.

In order to accommodate the particularist worry, our formulation of supervenience has to satisfy the following three desiderata:

[65]As well as, perhaps, the property of there not being any more potentially enabling/disabling/attenuating/intensifying properties to include.

1. The two normatively indiscernible individuals must be *intrinsically* non-normatively indiscernible;
2. The *worlds* of the normatively indiscernible individuals must be non-normatively indiscernible; and
3. The two normatively indiscernible individuals must be *extrinsically* indiscernible.

Call any two individuals that meet all three desiderata *maximally non-normatively indiscernible* or *fully non-normatively indiscernible*.

We need (1) because intrinsic properties can be relevant to the determination of normative properties. Whether someone is beautiful, for example, can vary with variations in that person's intrinsic relational properties such as "having symmetrical features". So we need to hold the intrinsic properties fixed. We need (2) in order to make sure that there are no differences between the worlds of the two individuals such that one world will provide an enabler, disabler, attenuator or intensifier that the other does not. We need (3) in order to make sure that the individuals are identically situated. Imagine an intrinsic duplicate of Hitler in our world that has the same intrinsic dispositions and characteristics and so on, but lives his whole life plugged into a computer simulation of early 20th century Germany. Although Hitler and his intrinsic duplicate are intrinsically non-normatively indiscernible, the real-life Hitler plausibly has certain moral properties that the intrinsic Hitler duplicate lacks—for instance, the property of being such that assassinating him is morally justified. This is because Hitler and the Hitler duplicate do differ in extrinsic properties: the Hitler duplicate has the property of being plugged into a computer simulation the real-life Hitler does not. By specifying that the individuals are also extrinsically indiscernible, we rule out

these sorts of cases from our formulation of supervenience.

It is likely that the three desiderata I give are redundant. For example, if two individuals are both intrinsically and extrinsically non-normatively indiscernible, then it seems likely that this would entail that they were in non-normatively indiscernible worlds. This is because among the extrinsic properties of individuals will be properties like "such that there is a White House at 1600 Pennsylvania Avenue, such that electrons have negative charge" and so on, and you can likely give an exhaustive non-normative description of a world in terms of extrinsic properties like these. Furthermore, if you think that the intrinsic/extrinsic property distinction is incoherent, then (1) and (3) can just be collapsed into the requirement that the individuals are non-normatively indiscernible simpliciter. I use the intrinsic/extrinsic distinction here because I think it better clarifies the argumentative structure of this part of the dissertation.

When we incorporate this into a weak supervenience formulation, we get what I'll call *Full* Weak Individual Additive Supervenience:

> *Full Weak Individual Additive Supervenience*: For any two possible worlds w_1 and w_2 and any individuals i_1 in w_1 and i_2 in w_2, if $w_1 \approx w_2$ and i_1 and i_2 are intrinsically and extrinsically non-normatively indiscernible, then, for any normative property P, $P(i_1)$ if and only if $P(i_2)$.

This is a formulation of weak supervenience because it specifies that the two worlds in question are the same world. I formulate it this way because it will be more easily allow us to revise this into a formulation of strong supervenience after I explain in the next few paragraphs why Full Weak Individual Additive Supervenience is problematic.

Full Weak Individual Additive Supervenience turns out to be either trivial or vacuous.

Take the case in which i_1 and i_2 denote the same individual. The conditional would then be trivially true. If i_1 and i_2 were to pick out the same individual in the same world, then they would trivially refer to individuals that were fully non-normatively indiscernible and also normatively indiscernible in that world.[66]

If $w_1 \approx w_2$, then i_1 and i_2 cannot denote numerically distinct individuals. Then the antecedent would never be satisfied, because no two numerically distinct individuals in the same world will ever be non-normatively indiscernible with respect to both their intrinsic and their extrinsic properties. Two non-normatively identical paintings, for instance, will at least be located in a different region of spacetime, so they will fail to be fully non-normatively indiscernible. You could in principle have a world that contained two indiscernible but numerically distinct individuals. There are of course the thought experiments from the identity of indiscernibles—for instance, Max Black's description of a world containing nothing but two inert iron spheres (Black 1952, p. 156). These examples, however, don't help us in our quest to formulate normative supervenience, since these sorts of examples always involve specifying highly artificial states of affairs that don't contain people, actions, or any of the kinds of things that have moral properties. In worlds with instantiated normative properties—and, of course, the one we most care about, ours—no two numerically distinct individuals will be both intrinsically and extrinsically non-normatively indiscernible. So the formulation turns out to be true but vacuous. This is a problem, since normative supervenience, whatever else it might be, is a substantive doctrine. People accept

[66]This follows from the claim that every property of an individual is either intrinsic or extrinsic; and Leibniz's indiscernibility of identity doctrine, which, when understood as distinct from his identity of indiscernibles doctrine, is, as far as I can tell, wholly uncontroversial. Denying it entails logical contradiction, since you'd be committed to saying that an individual both possesses and lacks a property.

normative supervenience for substantive reasons, not because they think it is only trivially or vacuously true.

Strong individual supervenience does not have this problem. Here is the relevantly modified version of strong normative supervenience:

> *Full Strong Individual Additive Supervenience*: For any two possible worlds w_1 and w_2 and any individuals i_1 in w_1 and i_2 in w_2, if w_1 and w_2 are non-normatively indiscernible and i_1 and i_2 are intrinsically and extrinsically non-normatively indiscernible, then, for any normative property P, $P(i_1)$ if and only if $P(i_2)$.

Slightly more colloquially, Full Strong Individual Additive Supervenience can be expressed as follows:

> Any two intrinsically and extrinsically non-normatively indiscernible individuals, situated in non-normatively indiscernible worlds, are normatively indiscernible.

Full Strong Individual Additive Supervenience replaces the condition from Full Weak Individual Additive Supervenience that the two worlds be identical with the condition that the worlds be non-normatively indiscernible. Although there may well be a problem with finding distinct, descriptively indiscernible individuals within a single world, there is no triviality/vacuity problem, at least notationally, in specifying two individuals in different worlds that are fully non-normatively indiscernible but differ with respect to their normative properties. It is still an open question at this point whether such a world is metaphysically possible, but we can at least talk about the world.[67]

[67] I assume that there is some way to engage in meaningful discussion about metaphysically impossible states of affairs—see, for instance, (Nolan 1997).

We had originally been concerned about strong individual normative supervenience because we wanted an ecumenical formulation of moral supervenience, and strong individual supervenience was denied by Hare and Dancy. However, once we make the move of limiting our formulation of strong normative supervenience to cover only only intrinsically and extrinsically non-normatively indiscernible individuals in non-normatively indiscernible worlds, much of the motivation for ethicists to support weak rather than strong supervenience evaporates. Hare, for instance, denied strong normative supervenience because he wanted to allow that non-normatively indiscernible actions could vary in whether we ought to do them with differences in our rational universalizable prescriptions. Fully non-normatively indiscernible individuals, however, can't vary in this way, because they could be distinguished by their property of being such that we rationally universally prescribe them. Similarly, extrinsically non-normatively indiscernible individuals in non-normatively indiscernible worlds cannot vary with respect to additional situational features that might change the valence or intensity of their moral features, because adding any such features would make the worlds non-normatively discernible.

For this reason, the formulation of supervenience that lays the best claim to being ecumenical is Full Strong Individual Additive Supervenience. This is this kind of supervenience formulation that I take to be uncontroversial not only among realists but among the anti-realists as well. It is, in other words, exactly what Dancy describes when he writes,

> It is better just to think of supervenience as... expressed in the fully general claim that if we start from a wrong action and move out to the entire non-moral nature of the world in which it is situated, and then replicate that in a new world, we are certain to have a wrong action in the replicating world. There is nothing more to supervenience than this. (Dancy 2004, p. 87)

For the reasons fundamentalist, all that is normative consists in facts about which reasons-apt tuples instantiate the favoring relation, and purely formal higher-order facts about these facts. The relevant application of Full Strong Individual Additive Supervenience for the formulation of reasons fundamentalism is therefore

> *Additive Reasons Supervenience*: For any possible worlds w_1 and w_2 and any fact-agent-action-circumstance tuples T_1 in w_1 and T_2 in w_2, if T_1 and T_2 are intrinsically and extrinsically non-normatively indiscernible and w_1 and w_2 are non-normatively indiscernible, then T_1 instantiates the favoring relation if and only if T_2 instantiates the favoring relation.

This is the formulation of supervenience that I will exploit in the arguments of §3 and §4.

Chapter 3

Irreducibility and Supervenience

In this chapter, I argue that the conjunction of Primitivism, Irreduciblism and Additive Reasons Supervenience presents an explanatory burden for the reasons fundamentalist that the reasons fundamentalist cannot answer. Call a defender of both Irreduciblism and Primitivism an *irreduciblist primitivist*. In this chapter, I argue that the irreduciblist primitivist either needs to give up Additive Reasons Supervenience (hereafter: supervenience unless otherwise specified), or accept that she cannot explain why Additive Reasons Supervenience expresses a metaphysical necessity. The failure to explain why Additive Reasons Supervenience expresses a metaphysical necessity threatens to generate for the irreduciblist primitivist the worries about denying supervenience that I discussed in chapter 2.

3.1 Explaining supervenience

The ban on mixed worlds

In the earlier chapter, in the section on supervenience, I discussed Zangwill's concept of a "Because Constraint" on normative properties: that normative properties can be instantiated in an individual only if there is some non-normative feature of that individual in virtue of which the individual instantiates the normative property. I argued that this was

not only plausible, but that its truth also entails a normative supervenience claim. This, in conjunction with the intuitive evidence and the worries about moral arbitrariness, provides good reason to accept a normative supervenience claim of some sort. I have argued that the kind of supervenience claim that we should believe on this basis is a strong individual supervenience claim.

As philosophers, though, we will not be content simply with accepting a normative supervenience claim on the basis of good reasons to accept it. There's a further question: why do normative and non-normative properties relate in this way? The Because Constraint, the intuitive support for supervenience, and the worries about moral arbitrariness provide good reasons to accept a normative supervenience claim, but they don't tell us why properties of these types relate in this way.

To be sure, the Because Constraint does say something about why normative properties should supervene on non-normative properties. Zangwill argued that the Because Constraint is both a conceptually and metaphysically necessary feature of normative concepts and properties. For Zangwill, it doesn't make sense to say that a property is a normative property unless it is instantiated in an individual, when it is instantiated in that individual, because of non-normative properties that that individual possesses, and it's also metaphysically impossible for normative properties to be instantiated unless they are instantiated in virtue of non-normative properties that the individual has. However, this sort of explanation simply pushes the demand for explanation back a further step. We can grant all these points that Zangwill wants us to grant, yet still wonder what sort of metaethical theory of normative properties best accounts for these features that normative properties possess. In other words, a metaethical theory that subscribes to normative supervenience should also

be able to say something about what it is about normative properties that gives rise to their supervenience on non-normative properties.

There is a long-standing concern, first presented by Simon Blackburn (Blackburn 1993), that the irreduciblist primitivist is not equipped to supply such an explanation. Since reasons fundamentalists are committed to irreduciblism and primitivism, Blackburn's concern about irreduciblist primitivism is a forteriori a concern about reasons fundamentalism.

Blackburn claims that the irreduciblist primitivist cannot appeal to identity or analytic entailment in her characterization of the relation between normative and non-normative facts. Irreduciblists cannot appeal to identity, because they deny that normative properties are identical to non-normative properties. Primitivists cannot appeal to analytic entailment, because primitivists deny that there is any conceptual analysis of normative properties. Without identity or analytic entailment, Blackburn asserts that the irreduciblist primitivist cannot explain why there should be such a restriction on the space of possible worlds—or a "ban on mixed worlds", as Blackburn likes to put it—as normative supervenience is.

Normative supervenience restricts the space of possible worlds because, if it obtains, then the relation rules out the possibility of worlds that are indiscernible with respect to their non-normative features but discernible with respect to their normative features. To illustrate, if in our world it is always and everywhere wrong to kill innocent people, the ban on mixed worlds states that no such world non-normatively indiscernible with ours can also be such that act utilitarianism is true.

Blackburn argues that the ban on mixed worlds requires an explanation that the irreduciblist primitivist cannot give. Blackburn's argument has exercised considerable influence over debates in metaethics, and, like the rest of Blackburn's important work in

metaethics, is worth serious attention.[1]

According to Blackburn, anyone who accepts normative supervenience is committed to the view that non-normative truths "entail" the normative truth. This poses a special problem for the irreduciblist primitivist, since the irreduciblist does not have resources to explain why the supervenience relation obtains. A similar concern is expressed by Mackie:

> What is the connection between the natural fact that an action is a piece of deliberate cruelty—say, causing pain just for fun—and the moral fact that it is wrong? It cannot be an entailment, a logical or semantic necessity. Yet it is not merely that the two features occur together. The wrongness must somehow be "consequential" or "supervenient"; it is wrong because it is a piece of deliberate cruelty. But what in the world is signified by this "because?" (Mackie 1977, p. 44)

Although supervenience relations are covariance relations, particular sets of facts rarely if ever *merely* covary: they usually do so for some reason. Facts about the existence of mereological sums supervene on facts about the existence of these sums' parts because the parts *compose* the sum; facts about the computational processes of circuits supervene on the electrical activity of these circuits because the electrical activity *realizes* the computational processes, and so forth. The mere fact of covariance is not philosophically satisfying: we want to know the how and the why.

In the quote from Mackie above, Mackie seems to mistakenly imply that the "because" in question here could only be explained by the presence of a logical or semantic necessity. The examples of parthood and realization show that there are more ways to explain supervenience than through an appeal to logical or semantic necessity. If the normative properties are realized by or composed of the non-normative ones, the presence of the normative-

[1]Michael Ridge extends Blackburn's argument in (Ridge 2007). See also (Olson 2012).

nonnormative realization or composition relation would explain normative supervenience. This is because, after all, realization and composition relations just *are* varieties of supervenience relations. In principle, the primitivist irreduciblist could appeal to more specific varieties of covariance in order to explain normative supervenience, thereby avoiding Mackie's concern. For example, Brink (Brink 1989) argues that normative properties are realized by non-normative properties. I myself do not think this approach is likely to be successful, and will argue against this approach in §4.3.2 below.

Blackburn calls his target the normative *realist*; however, the argument that he offers against reduciblist formulations of realism is not the argument I've just given. His concern with reduciblism is not that reduciblists lack an explanation of supervenience. Reduciblists can trivially explain the supervenience of normative properties on descriptive properties because reduciblists think that normative properties just are descriptive properties. Blackburn's objection to reduciblism is that the reduciblist can't explain why supervenience is not only a metaphysical constraint on normative property instantiation, but also a conceptual constraint. This is also a concern that Blackburn has with irreduciblist primitivism.

Irreduciblist responses

Irreduciblists who take issue with Blackburn's argument are disposed to bang their fist on the point that an individual's non-normative properties entail that individual's normative properties. Irreduciblists are fine with admitting that there is such an entailment. Moore (the paradigmatic irreduciblist), for example, states, "If a thing is good (in my sense), then that it is so *follows* from the fact that it possesses certain natural intrinsic properties, which

are such that from the fact that it is good it does not follow conversely that it has those properties" (Moore 1922, p. 261).

This sort of response has been developed by contemporary irreduciblist philosophers. Consider, first, Russ Shafer-Landau, who argues contra Blackburn that irreduciblists of his sort do have a ready explanation of why the normative supervenience relation obtains. Says Shafer-Landau:

> We can explain the ban on mixed worlds by claiming that a duly specified set of non-moral properties metaphysically must give rise to a certain moral property. The dependence relation referred to by Moore can take the form of an entailment relation specifying metaphysically sufficient non-ethical conditions for the instantiation of moral properties... Competent speakers can conceive of a world in which the base properties that actually underlie particular moral ones fail to do so. But there is no mystery here, since people can conceive of many things that are not metaphysically possible. (Shafer-Landau 2003, p. 86)

The key point to note here is Shafer-Landau's claim that the entailment relation that is supposed to explain normative supervenience is construed as a relation that obtains with metaphysical necessity.

This claim that the entailment between non-normative facts and normative facts obtains with metaphysical necessity seems to be fairly common among the irreduciblists who are interested in this question. The most radical, card-carrying contemporary irreduciblist, David Enoch, writes,

> What is the precise modal status of the basic moral norms? I am not sure what to say here. It should be at least metaphysical necessity, so that there is no world where the basic norms are different. (Enoch 2011, p. 146)

Scanlon's views in *Being Realistic About Reasons* seem to lean in this direction as well. He claims that the pure normative truths—things like "pain is ceterus paribus bad"—are in

fact necessary and are not contingent in any metaphysical sense (Scanlon 2014, p. 41). As such, the presence of any pain state will also "entail" the presence of badness in that state. Both Enoch and Scanlon here are asserting that there are fundamental normative principles such as "pain is bad" that are necessarily true as a matter of metaphysical necessity. There are no metaphysically possible worlds in which this principle is false. Take some mental state x that is a pain state. Because "pain is bad" is necessarily true, "pain is bad" will be true in the world featuring x. Therefore x will also be bad, and we have a metaphysical entailment from the fact that x is a pain state to the fact that x is bad.[2]

By "metaphysical necessity", what Shafer-Landau, Enoch and Scanlon have in mind is something like the following: something A entails something else B with metaphysically necessity just if God doesn't have to do anything other than to make A in order to make B. As Kripke says,

> Suppose we imagine God creating the world; what does He need to do to make the identity of heat and molecular motion obtain? Here it would seem that all He needs to do is to create the heat, that is, the molecular motion itself. If the air molecules of this earth were sufficiently agitated, if there is a burning fire, then the earth will be hot even if there were no observers to see it.... (Kripke 1980, p. 153).

To make heat, all God needs to do is to make the molecules move. To make cat-fusions, all God has to do is make the cats. To make it not the case that chalk is not white, all God has to do is make chalk white. No further divine effort need be expended. Similarly, if it's

[2]Scanlon, Enoch and Shafer-Landau have distinct views as to what reasons are and to why the normative necessities are metaphysically necessary. For a treatment of these differences, the best place to look is (Vayrynen 2016). I am not discussing these differences because I am focusing on the claim, common to all three irreduciblists, that the normative necessities are metaphysical necessities.

metaphysically necessary that pain be bad, all God needs to do to make pain states bad is to make the pain states. Their badness will come for free.[3]

Kripke's theistic metaphor of metaphysical necessity is just that: a metaphor. It's not meant to be an analysis of metaphysical necessity. There is no agreed-upon analysis of metaphysical necessity in the literature, so unfortunately I can't provide you with one. I am talking about the Kripkean metaphor here just to give you a heuristic that you can use to pump your intuitions about what entailment relations would count as metaphysically necessary.[4]

Some philosophers argue that metaphysical necessity is a confused notion. In recent work, Justin Clarke-Doane argues that there is no sense of metaphysical modality that has the kind of privileged status that NYU-style philosophers grant to it. Asking whether something is metaphysically possible, for Clarke-Doane, is like asking whether the Parallel Postulate is true: there is no absolute sense of the Parallel Postulate such that it is either true or false. What we rather say about the Parallel Postulate is that it is true of Euclidean geometries and false of non-Euclidean geometries. Similarly, whether some proposition counts as being "metaphysically" possible or necessary depends on what kind of restriction you've already drawn on the space of worlds.

Clarke-Doane suggests that what this means is that there is no interesting question as to what counts as metaphysically possible or necessary. The moral I draw from his argument is somewhat different. I think that what Clarke-Doane has argued shows not that there is

[3]Even if it were metaphysically necessary that pain be bad, this does not provide a solution to the theological problem of evil, since there still remains the question why God created pain or anything else that is bad as a matter of metaphysical necessity.

[4]I'm not the only one to do this, e.g. (Enoch 2011, p. 141) and (Chalmers 2002a, p. 146).

no interesting question as to what is metaphysically possible and metaphysically necessary, but rather that there are very many interesting questions as to what counts as metaphysically possible or metaphysically necessary. The sense of whether normative necessities are metaphysical necessities that interests me in this dissertation is the sense in which Enoch, Shafer-Landau and Scanlon mean it: whether, once one has specified all the non-normative facts in a world, the normative facts that also exist in that world come for free.

To repeat, Shafer-Landau, Enoch and Scanlon are all irreduciblists who assert that the non-normative facts "entail" the normative facts, and that the entailment in question should be understood not as conceptual or analytic entailment but as metaphysical entailment, or entailment with metaphysical necessity. Unfortunately, this kind of response misses the point of Blackburn's argument. Let's grant with Enoch and Scanlon that the non-normative facts necessitate the normative facts, *given* that the normative principles or pure normative truths are as they are. This would be to restrict the space of worlds to those worlds in which the pure normative truths are as they are. I'll follow Kit Fine in calling this kind of necessity *normative necessity*. By itself, however, normative necessity does not immediately give us the *metaphysical necessity* that Enoch, Shafer-Landau and Scanlon want to impute to the normative supervenience claim. We can imagine Max, a fan of Blackburn, asking something like this: "I admit that the non-normative features of the world give rise to the normative features that it does as a matter of normative necessity. However, I deny that the non-normative features of the world give rise to the normative features that it does as a matter of metaphysical necessity. For I deny that the normative necessities are metaphysical necessities. Why should I think otherwise?" Although pure normative truths such as "pain is bad" may, as Shafer-Landau argues, give us an entailment relation from pain

states to badness, simply claiming that this entailment relation obtains with metaphysical necessity does nothing to answer Max's question. The lack of an answer to this question is intolerable.

Perhaps metaphysical necessity in this case just comes for free. After all, metaphysical necessity does come for free sometimes. Lewis, for example, argued that the mereological facts come for free once one has assembled all facts about the mereological sums' compositional bases. Humeans about physical laws argue that the physical laws come for free once one has the Humean mosaic.

However, this sort of response won't work for the irreduciblist. The reason why fusions come for free with their parts is because, as Lewis says, "The fusion [of cats] is nothing over and above the cats that compose it. It just *is* them. They just *are* it" (Lewis 1991, p. 81). The reason why physical laws come for free is because laws, for the Humean, just are the exceptionless constant conjunctions. Once you have the Humean mosaic, you have all the exceptionless constant conjunctions and therefore all the laws.[5] The irreduciblist, however, specifically holds that the normative properties *do not reduce*: they are indeed something over and above the non-normative properties that give rise to them. So the question arises: why are the normative necessities metaphysical necessities? How can they come for free when, for the irreduciblist, the normative facts don't simply reduce to the non-normative facts?[6] Shafer-Landau asserts, true enough, that the non-normative/normative relations can be metaphysically necessary even if they are not conceptually necessary. Of course. This,

[5]In saying this, I am not claiming that Hume himself held the Humean view of physical laws. Whether Hume held this view is a difficult interpretive issue—see (Beebee 2006, pp. 108-141).

[6]Again, the irreduciblist can appeal to realization to offer an explanation here. I criticize this sort of response in §3.3.2.

however, is not in itself a reason to think that they *are* metaphysically necessary, but only that they *may* be metaphysically necessary even if there is no conceptual entailment between a full description of the non-normative features of a world and a description of the normative features of that world. It seems clear, then, that, in this exchange, Blackburn has the better hand.

A proponent of the claim that the normative necessities are metaphysical necessities might simply put her foot down and say that the normative necessities are metaphysical necessities, full stop. I take it, though, that in debates about what is metaphysically necessary, the burden of proof falls on the affirmer. A claim that some description is true of a world as a matter of metaphysical necessity is ipso facto a claim that other descriptions of some world that are inconsistent with the first are necessarily false. If one claims, for instance, that it's impossible that there be no homogenous, enriched uranium spheres that are one mile in diameter, we want to know what is wrong with these spheres that makes them impossible. After all, there's no problem with homogenous lead spheres of that size, as far as we can tell, and we have no problem accepting the metaphysical possibility of these lead spheres. We assume that a description is a description of a metaphysically possible world until such time that someone gives us a reason to think otherwise: say, formulates the description in a way sufficiently precise to the details, and then points out some inner inconsistency or incoherence in the concepts involved in the fuller description. As it turns out, spelling out the notion of a mile-diameter enriched uranium sphere in further detail would reveal that the alpha decay of a mass of enriched uranium even a fraction of that size would trigger a chain reaction that would both fragmentize the mass and convert the enriched uranium into something else. But in the absence of this sort of explanation and precisification of the

details of the description, we tend to default to thinking that any given description describes a possible rather than an impossible state of affairs. So it seems as though defenders of the claim that the normative necessities are also metaphysical necessities need to do more than stick to their guns. They do owe us an explanation—or at least an explanation as to why they don't owe us an explanation.

3.2 The Supervenience Dilemma

The Dilemma formulated

Here is a more concise statement of the problem that supervenience poses for the irreducibilist primitivist:

1. *Supervenience Disjunction*: Additive Reasons Supervenience expresses a metaphysical necessity, or it does not.
2. *Brute Entailment*: If Additive Reasons Supervenience expresses a metaphysical necessity, then the reasons fundamentalist can't explain why.
3. *Brute Normativity*: If Additive Reasons Supervenience does not express a metaphysical necessity, then reasons fundamentalists must either accept that there is a non-additive formulation of supervenience that expresses a metaphysical necessity, or reasons fundamentalists must accept that normativity is metaphysically arbitrary.
4. *Additive Exclusion*: The only variety of supervenience that reasons fundamentalists can accept is Additive Reasons Supervenience.
5. Therefore, either the reasons fundamentalist cannot explain why Additive Reasons

Supervenience expresses a metaphysical necessity, or reasons fundamentalists must accept that normativity is metaphysically arbitrary.

Call this the *Supervenience Dilemma* for the irreduciblist primitivist.

Here's a synopsis of the argument. The reason this is a *Supervenience* Dilemma for the irreduciblist is because supervenience relations are entailment relations. If B supervenes on A, then whenever you have A, you have B. A entails B. It's a *Dilemma* because either disjunct is bad for the reasons fundamentalist. Supervenience Disjunction is true because it's a logical tautology. Brute Entailment encapsulates the worry expressed in §3.1—namely, that any irreduciblist primitivist does not have an explanation of why Additive Reasons Supervenience expresses a metaphysical necessity. I believe that this is true, but will consider this claim in substantially more detail in §3.3 below. Additive Exclusion is the conclusion from the introduction to §2.5. In that section, I argued that Centrality commits any reasons fundamentalist that accepts normative supervenience to Additive Reasons Supervenience, and furthermore that there is no other, non-trivial normative supervenience claim that they can accept. Brute Normativity expresses the arbitrariness worries from §2.6.2 above—namely, that if the normative laws were not metaphysically necessary, then they could have been different than they are, in which case the fact that our world has the normative laws that it does seems un-moored ("un-Moored"...?) from the non-normative features of our world in a way that seems inexplicable and unsettling. The conclusion is that either that reasons fundamentalists can't explain why Additive Reasons Supervenience expresses a metaphysical necessity, or that they are committed to the metaphysical arbitrariness of normativity.

Note that the conclusion of the Supervenience Dilemma is only bad for the reasons

fundamentalist. Critics of reasons fundamentalism can easily accept the conclusion, since they can easily accept that their reasons fundamentalist opponents must choose between one of the two unpalatable disjuncts expressed in the conclusion. Too bad for them.

However, for the reasons fundamentalist, the conclusion of the Supervenience Dilemma is quite bad indeed. Every prominent metaethical theory (and probably the non-prominent ones too) has major problems: the reduciblist has difficulty explaining open question intuitions and modal reference conditions for moral terms, error theories are quite revisionary, noncognitivists face the Frege-Geach problem and Andy Egan's concern that they are committed to "moral smugness" (Egan 2007), and so on. But what they do promise to offer, at least, is an answer to the question what normative reality is and why it is the way it is rather than some other way (or is not, for the error theorist). These two questions are arguably the two defining questions of metaethical inquiry.

Irreduciblist primitivists have historically struggled with answering the first question—the very name *ir*reduciblism betrays this fact. A straightforward reading of the first chapters of the *Principia Ethica* suggests that Moore is arguing that, if someone asks us what goodness *is* (as opposed to asking us *what things are good*), there is nothing we can positively say. Moore gives us an unsatisfying *via negativa* for metaethical knowledge: we can't truly say anything about what goodness is, and can only truly say what it is not. If irreduciblists can't overcome the Supervenience Dilemma either, then it seems that irreduciblism cannot answer the second defining question (why normative reality is as it is rather than some other way), significantly lowering its prospects against the other views despite their own deep challenges. Those who find these worries unsettling can of course just try to look harder for a way to avoid *Brute Entailment*. We can do this, and we should. I will in fact

attempt to do this in §3.3. The worry, however, is that we may find the explanation we are searching for only by defecting from the irreduciblist primitivist camp.

Defending the Dilemma

Before considering the premises of the Supervenience Dilemma, I want to discuss objections that one might want to raise to the argument itself. Two objections in particular come to mind. One objection is that one or more of the premises of the argument are incoherent, which I'll consider first. Another objection, which I'll also consider, is that the premises of the argument are coherent, but the argument itself is invalid.

Is non-normative/normative entailment coherent?

I mentioned in §2.6.1 that there are only two philosophers that explicitly fail to commit themselves to the existence of a normative supervenience constraint on the distribution of normative properties. The first is Rosen (Rosen 2002), who thinks that neo-essentialism about modality gives us reason to think that normative supervenience is false. I discuss Rosen's concerns extensively in §3.3.3. The second is Bilgrami (Bilgrami 2006), who argues that we do not have a coherent concept of entailment relations between non-normative properties and normative properties. I discuss Bilgrami's argument in this section.

Bilgrami discusses normative supervenience in the context of an involved argument for what Bilgrami calls *perspectival duality*. Perspectival duality (which is not, of course, to be confused with substance dualism, property dualism, or dualism of other kinds) is *perspectival* because it is a duality of distinct *points of view*: the first-person point of view of agency and engagement within which norms and values are in play, and the third-person

point of view of "detachment" from these norms. It is constitutive of practical agents that they occupy the first-person point of view.

The *duality* of the perspectival duality refers to an interesting property of these points of view: that they are such that it is impossible for an individual to occupy both of them at once. Features of the world that are conceivable while occupying the agential perspective (call these "agential features") are not conceivable while occupying the detached perspective, and features of the world that are conceivable while occupying the detached perspective (call these "detached features") are not conceivable while occupying the agential perspective.[7]

Bilgrami furthermore argues that there is no third, "straddling" perspective from which one can conceptualize both the agential features of the world and the detached features of the world at the same time. Occupying one point of view requires a switch in perspectives such that the various features of the world that would have been present when occupying the previous point of view are no longer able to be accessed and evaluated. Because we cannot conceptualize agential features and detached features at the same time, we have no way of conceptualizing relations between agential and detached features.

The result, Bilgrami argues, is that normative supervenience relations are incoherent. Normative supervenience is a relation that obtains between the normative features of the world and the non-normative features of the world. However, normative features of the world are agential features: they are conceivable only while occupying the agential perspective. Non-normative features of the world are detached features: they are conceivable

[7] The phrases "agential features" and "detached features" are mine. To my knowledge, Bilgrami does not use these exact phrases.

only while occupying the detached perspective. Normative supervenience is therefore a relation that obtains between agential features of the world and detached features of the world. But, as we just saw, it is not possible for us to conceive of relations between agential and detached features of the world, since this would require us to occupy a third point of view that straddles these two perspectives, and such a point of view does not exist. Therefore normative supervenience is inconceivable and ultimately incoherent.

Explicit criticism of Bilgrami's supervenience claim has focused not as much on the existence of the agential and detached perspectives themselves, but on the claim that relations between them cannot be accessed. Baldwin and Normore, for instance, both suggest that the sorts of facts that we access and evaluate when occupying the detached point of view are precisely the ones that feature in practical reasoning, and therefore that agents are capable of understanding and evaluating normative supervenience theses while occupying the agential point of view (Baldwin 2010), (Normore 2010). Even if there exists no perspective that unites the agential and detached perspectives, Baldwin and Normore suggest that the features of the world available to us while occupying the detached perspective are also available to us while occupying the agential perspective. As Normore writes,

> From your perspective you see me as having (say) needs in virtue of my low caloric intake. You can indeed see the caloric intake itself as calling on you—it's too low—less than required to keep me in good health—and you conclude something must be done about it. You plan to bring me meals on wheels. You are five kilometers away. You are 15 minutes away as you drive and ten minutes away as your partner drives. My low caloric intake is partly a product of circumstance and partly of a genetic disorder. There is nothing about the third person perspective that you need to leave out in any of this. (ibid., p. 768)

Baldwin and Normore construe the detached perspective as an impoverished version of the

agential perspective. Although we can occupy the detached perspective and thereby lose access to the normative features of the world, simply switching back to the agential perspective brings back our access to these features without losing anything we were able to access while occupying the detached perspective. Switching from one perspective to the other is like putting on your normative sunglasses and then taking them off again. Now the world is laden with normativity; now it's not; that's the only difference and all there is to say. If all the features of the world available to us while occupying the detached perspective are also available to us while occupying the agential perspective, then perspectival duality gives us no reason to think that normative supervenience is incoherent. We can easily conceive of the normative supervenience relation and evaluate it while occupying the agential perspective, from which we can access both the detached and agential features of the world.

There does seem to be something prima facie plausible about Normore's suggestion. Surely we take facts about the caloric properties of certain substances into consideration when preparing meals to victims of diabetes. However, I take it that the objection misses the distinctive and radical feature of Bilgrami's view. Bilgrami's suggestion, I take it, is not simply that occupying the agential point of view involves value judgment or even that value-judgments are only understandable from the agential perspective. Bilgrami's radical suggestion is that occupying the agential point of view—which is to say, being an agent at all—involves understanding the world as being "saturated" through and through with value and normativity. As Bilgrami writes,

> It makes all the difference that [in the agential point of view] I am describing the person or persons as being *in need*. To do so is to observe a fact that is *laden with value* in the sense that I perceive it to be making certain *evaluative or normative or imperatival demands on me*... The first person point of view...

> sees the world in a way that is laden with value–in a way, therefore, that implicates the perceiving agent, involves the perceiving agent's engagement with it. (Bilgrami 2006, pp. 256-257)

For Bilgrami, values and norms are not "additive" properties that sit atop a barren world that is, in itself, "polar night of icy darkness"—a description Weber used to characterize societies organized largely in accord with the demands of bureaucratic, means-ends instrumental rationality. There are no normative sunglasses that you can put on and take off at your leisure. The world's saturation with value is such that, once an agent ropes together all the value-laden properties in order to formulate a supervenience claim, the set of properties that remain would not be sufficiently robust to form a supervenience base. Perhaps there would not be any properties to rope together at all. The opposite problem harasses attempts to formulate normative supervenience from outside the agential perspective. From the detached, non-agential perspective, nothing is normative and nothing is value-laden. Once such an individual ropes together all the non-normative features of the world (if such a thing is even possible), there would be no properties left to play the supervening role. Baldwin considers this sort of response but dismisses it as "absurd". Calling a view absurd, however, does not amount to an argument against that view.

I accept Bilgrami's claim that there exists a perspectival duality between the first-person and third-person points of view. I also accept that, since these points of view cannot be occupied at the same time, claims about relations that hold between facts accessible only from one perspective, and facts accessible only from the other perspective, are unintelligible and therefore cannot be accessed. I furthermore accept that, when suitably spelled out, the normative features of the world are accessible only from the agential perspective, and the

non-normative features of the world are accessible only from the detached perspective. I therefore accept that any supervenience doctrine that purports to express a relation between normative and non-normative properties is unintelligible.

Additive formulations of supervenience, as I have described them, purport to express a relation between normative and non-normative properties. I therefore conclude that additive formulations of supervenience are unintelligible. However, the arguments in this dissertation against additive supervenience will not draw on Bilgrami's arguments. I am not drawing on Bilgrami's arguments simply because many philosophers don't accept perspectival duality, and I want to convince them too. Fortunately for me, there are good reasons even for critics of perspectival duality to fail to accept additive supervenience. I give these reasons in §5.3. In what follows, then, I will leave perspectival duality behind and argue as though I and other people understood what an additive supervenience relation could be.[8]

Secondly, I will argue that Bilgrami's arguments do not establish that normative supervenience *in general* is incoherent. It establishes only that *additive* formulations of normative supervenience are incoherent, but leaves intact the coherence of supervenience relations of other, non-additive kinds. In particular, I argue that it is coherent on the perspectival duality picture to say that normative properties supervene on descriptive properties, provided that these descriptive properties are also normative in some sense. I will exploit this idea when developing and defending my own, non-additive ("transformative") formulation of

[8]This strategy of arguing against views one takes as incoherent has a long tradition in philosophy. Although Anselm thinks it's incoherent to say that God does not exist, he still countenances the fool who sayeth in his heart there is no God. Although Hume believes that philosophers do not have an idea of necessary connection, he still argues with them as though he did. And so on.

normative supervenience in §5.2 and §5.3.

Revising logic

Perhaps irreduciblist primitivists can avoid the Supervenience Dilemma by claiming that the Supervenience Dilemma is invalid. The Dilemma is valid, of course, but in classical logic, and irreduciblist primitivists can argue that classical validity is not the kind of validity we are looking for when evaluating the validity of arguments such as the Supervenience Dilemma. Perhaps we have reasons to think that classical logic is not the logic that best captures the kind of reasoning at work when we consider the phenomena the argument discusses.

Some phenomena don't easily fit into classical logical models. Vagueness and the semantic paradoxes are two good examples. Michel Foucault is definitely bald, and David Chalmers is definitely not. What about Kit Fine? It's not clear. One response to this lack of clarity about how to apply vague predicates in borderline cases is to say that people on the borderline of baldness are neither bald nor not bald, but rather indeterminately bald. The introduction of indeterminate truth values entails a denial of the Law of Excluded Middle and marks a departure from the semantics of classical logic.

Similarly, the sentence "This sentence is false" seems to be both true and false. If the sentence is true, then it is false; if it is false, then it is true. One response to this apparent overdetermination of truth values is to say that the sentence is neither "true and not false" nor "false and not true" but rather "both true and false". The introduction of overdetermined truth values entails a denial of the Law of Bivalence and marks a departure from the semantics of classical logic.

Irreduciblists cannot avoid the Supervenience Dilemma by denying the Law of Bivalence—say, by claiming both that the normative necessities are metaphysical necessities and that the normative necessities are not metaphysical necessities. This actually makes the case worse for them, since they would then be committed not just to one of the bad consequences expressed in *Brute Entailment* and *Brute Normativity*, but rather both of them.

Irreduciblists could deny the Law of Excluded Middle. This would allow them to claim that it's indeterminate whether the normative necessities are metaphysical necessities, which would allow them to navigate the horns of dilemma expressed in the argument's conclusion. I take it, though, that the rejection of classical disjunction, like claims that any given description is in fact impossible, would amount to a rejection of a default view. Just as we should hold that descriptions are default possible until given good reason to think otherwise, we should hold that truth-apt sentences, like the disjuncts in Supervenience Disjunction, are default *bivalence-conforming*: such that they are either determinately true and determinately not false, or determinately false and determinately not true. Revisions to logic are very big deals indeed (e.g., (Williamson 1996)).

Perhaps the phenomena at stake in the Supervenience Dilemma are important enough to motivate a rejection of classical logic. To be sure, there likely are some phenomena in the philosophical universe that do require deviation from classical logic in order to make sense of them, and too important to just give up entirely. Normativity, in my view, is to important to give up. So, if normativity were one of these phenomena that required a deviation from classical logic in order to make sense of it, then we should deviate from classical logic.

However, what's at stake in Supervenience Dilemma is not normativity. It is irreducib-

lism. And although irreduciblism has many virtues (I am an irreduciblist, and I wouldn't be one if I didn't think it had virtues), it is not so important that we should sacrifice classical logic for it. If we have to choose between irreduciblism or classical logic, we should go for classical logic, no questions asked.

Furthermore, even if, for whatever reason, irreduciblism were important enough to sacrifice classical logic, we should sacrifice classical logic for it only if there were no other way to save irreduciblism. This dissertation will argue in §5 that there is a way to save irreduciblism that does not involve any revisions to classical logic. So we shouldn't revise our logic in order to save irreduciblism.

Chapter 4

Avoiding Brute Entailment

Can irreduciblist primitivists reject Brute Entailment? In §3.1, I raised a problem for irreduciblist primitivist's capacity to explain why the non-normative features of our world necessarily give rise to the normative features of the world that they do. I argued in §3.1 and §3.2 that the irreduciblist primitivist cannot explain this within the resources provided by their account of normativity. However, it's worth investigating whether they can supply such an account on the basis of resources provided by modal epistemology or modal metaphysics. Specifically, the irreduciblist primitivists can attempt to find an independent account of metaphysical modality such that this account provides an explanation of why Additive Reasons Supervenience expresses a metaphysical necessity rather than simply a normative necessity.

In this section, I'll consider three distinct approaches to construing metaphysical necessity and investigate whether they can help the irreduciblist reject Brute Entailment. The three approaches to metaphysical necessity that I'll consider are conceivability approaches, specifically, the conceivability approach of Chalmers; causal regulation approaches, specifically, the approach of Brink; and neo-essentialist approaches, specifically, the approach of Kit Fine.

There are of course many more characterizations of metaphysical necessity than the

ones that I'll consider. In particular, I am considering in detail neither David Lewis' modal realism nor David Armstrong's combinatorialism. About modal realism I have little to say, finding it highly implausible. Armstrong has an ontology of simple particulars, simple universals and simple relations; his combinatorialism about modality is the view that all the combinations of simple particulars, properties and relations that respect certain formal constraints constitute possibilities (Armstrong 1997, p. 160).[1] On the recombination principle, a particular combined with a simple normative property and relation like the reason relation will be a possibility, and a particular combined with the absence of this property or relation would constitute another possibility. So Armstrong's combinatorialism would not generate generate metaphysical necessity in the normative case.

I choose the conceivability, a posteriori and neo-essentialist approaches to metaphysical possibility and necessity because of their current particular prominence in the literature on modal epistemology and modal metaphysics. I will argue in this section that none of these three approaches ultimately gives irreduciblist primitivists sufficient resources to explain why Additive Reasons Supervenience expresses a metaphysical necessity.

4.1 Conceivability

Notions of necessity are closely linked to notions of possibility: a proposition P is necessary just if $\neg P$ is not possible. An investigation into metaphysical possibility is therefore eo ipso an investigation into metaphysical necessity.

Historically, the most common bridge to metaphysical possibility has been through con-

[1] A simple particular, property or relation is a particular, property or relation that cannot be further decomposed into constituent particulars, properties or relations.

ceivability. Conceivability-based arguments underlie many of the most famous arguments in philosophy, from Anselm's ontological argument to Descartes' argument for the numerical distinctness of mind and body. Notable examples in the 20th century include Moore's open question argument and Gettier's argument against justified true belief analyses of knowledge. Conceivability approaches faced a considerable challenge in the twentieth century from Putnam, Burge and Kripke's discovery of necessary a posteriori metaphysical necessities. But the conceivability approach also enjoys prominent defenders. In this subsection, I will interact with the conceivability approach with the help of Chalmers' helpful distinctions between varieties of conceivability.

Chalmers defends a variety of modal rationalism. For Chalmers, there is a characterization of conceivability such that all worlds conceivable in this way are possible. In characterizing the kind of conceivability at stake in his modal rationalism claim, Chalmers draws three binary distinctions across the conceivability space: ideal vs prima facie conceivability, positive vs negative conceivability, and primary vs secondary conceivability. The resulting framework therefore gives us a logical space of eight kinds of conceivability. I will first introduce these three distinctions, then discuss their implications for metaphysical necessity, and end with a consideration of whether Chalmers' approach can help the irreduciblist explain why additive supervenience is metaphysically necessary.

Types of conceivability

Ideal vs prima facie conceivability

It is prima facie conceivable that $18593 \times 4992 = 92716256$. (It doesn't, by the way.) At first glance, maybe that's the right answer. I am relatively good at arithmetic, but I am not a computer (although, if Chalmers's views on the singularity and personal identity are correct, perhaps I might become one in the not-too-distant future). As it turns out, ideal reflection tells me that that's not the right answer. In fact, such reflection tells me that it *could not be* the right answer.[2] The prima facie conceivability of "$18593 \times 4992 = 92716256$" is therefore not a very reliable guide to what is truly possible. An appropriate specification of the link between conceivability and possibility will then require that the conceivability be *ideal* conceivability: conceivability on ideal rational reflection. Specifically, a hypothesis S is ideally conceivable just if ideal rational reflection detects *no contradiction* in S; or, equivalently, when $\neg S$ is not ruled out a priori after ideal rational reflection.

Positive vs negative conceivability

We can *positively* imagine flying horses and means of production that are democratically controlled. These are things that we can coherently *imagine*, or picture in our mind's eye if you will. We can also imagine invisible creatures, dark matter, angels, ectoplasm and so on. These are things that we can't picture in our mind's eye (or at least, not without cheating), but that we can nevertheless coherently imagine. Chalmers lumps both varieties of imag-

[2]For the extreme minority view according to which the truths of mathematics are metaphysically contingent, see (Rosen 2002). Ayer of course worries that Mill's theory makes the arithmetical truths contingent: (Ayer 1952, p. 74).

ination under the label *modal imagination*, which he leaves as an intuitive notion that he thinks we all can grasp. I will follow him in this. Positively conceivable hypotheses are thus hypotheses that we can coherently modally imagine, where coherently modally imaginable hypotheses are hypotheses that we can either picture in our mind's eye (flying horses), or that we can imagine in the way that we imagine dark matter and invisible creatures.

Negatively conceivable worlds are worlds that we can't coherently modally imagine, but are nonetheless not ruled out a priori. In other words, a negatively conceivable world W is one such that there is *no apparent contradiction* in W. Since positively conceivable worlds are only ones that we can coherently imagine, positively conceivable worlds will therefore also be negatively conceivable on this definition.

There is a clear link between negatively conceivable hypotheses and ideally conceivable hypotheses. An ideally conceivable hypothesis is a hypothesis such that ideal rational reflection does not reveal any contradiction. A negatively conceivable hypothesis is one such that reflection does not reveal an apparent contradiction. Since ideal rational reflection is a species of reflection, an ideally conceivable hypothesis will therefore also be a negatively conceivable hypothesis, on the definitions given here.

Primary and secondary conceivability

It is impossible to give an exposition of primary and secondary conceivability that's both adequate and concise. What I give here is as brief and clear of a presentation as I can, in the context of a generic two-dimensionalist (2D) picture. Other introductions can be found at (Chalmers 2004), (LaPorte 2006), and (Soames 2004).

Two-dimensionalism is a semantic theory that aims to reconcile aspects of Fregean se-

mantics with the revolution in semantics launched by Kripke, Putnam, Burge and Devitt's discovery of, and investigation into, a posteriori necessities. I'll start with a very slightly technical description of the two-dimensionalist view and then move on to a more intuitive characterization.

In its broadest formulation, two-dimensionalism is a variety of possible-worlds semantic theory on which token utterances in a world are evaluated relative to not one but two intensions (or "dimensions"), which we can call their primary and secondary intensions, or 1- and 2-intensions. Both primary and secondary intensions are functions from possibilities to extensions. We can use (along with Chalmers) the term "scenario" to denote the possibilities salient to evaluation relative to primary intensions, and the term "world" to denote the possibilities salient to evaluation relative to secondary intensions. In addition, each strong assertion is associated with a two-dimensional intension, which is a function from {scenario, world} pairs to extensions.

On all two-dimensional frameworks, evaluation of a token utterance relative to its secondary intension is the standard Kripkean evaluation from worlds to extensions. A token utterance of "water" has a secondary intension that takes a world as input and gives us an extension consisting of all and only the H_2O in that world. We can call this "secondary evaluation" or "evaluation considered as counterfactual". Chalmers calls it "evaluation considered as counterfactual" because when we, say, consider the Twin Earth/XYZ world, we are considering whether the Twin Earth world represents some other way the world could have been. In actuality, the actual world is one in which H_2O causally regulates our use of "water", and thus the rigidification of the term gives us an empty extension in the XYZ world when we evaluate the XYZ world as counterfactual.

The first dimension of evaluation is more complicated, and varies from two-dimensionalist to two-dimensionalist. Loosely speaking, however, the first method of evaluation will be evaluation relative to *epistemic scenarios* specified in such a way as to capture aspects of a Fregean view of meaning, where the scenario is elucidated in different ways depending on the theory and the variety of phenomenon for which the 2D analysis is being offered: as Kaplanian characters, Stalnakerian propositional content, descriptions, narrow content, modes of presentation, centered worlds, and so on. We can call the method of evaluation relative to intensions of this kind "primary evaluation" or "evaluation considered as actual".

Here is an example. You know lots of things: things about reptilian biology, physics, metaphysics, everyday facts, and so on. You receive a small package in the mail. Before you open this box, your epistemic scenario will include all the things you knew before you received the package, plus the fact that you just got a package in the mail. For all you know, it might contain a diamond necklace, a box of chocolates, a book on the metaphysics of modality, or nothing at all—these are all ways that the world could be, given the epistemic scenario. In other words, if you consider a possible world in which the box contains a diamond necklace, nothing about that possible world conflicts with anything you currently know about the actual world. You do know, however, that the package does not contain dragon eggs, a supernova, or Sylvan's box. This is because there is something about each of the possible worlds in which that box contains dragon eggs, a supernova, or Sylvan's box that conflicts with what you currently know about the actual world—namely, the facts of reptilian biology, astrophysics, and metaphysics.

2D semantics uses these epistemic scenarios to explain our intuition that, if we were an

early 18th-century chemist, for all we know, water could turn out to be XYZ. For a chemist before 1750, water could turn out to be H_2O, or it could turn out to be XYZ (or both!)—her epistemic state doesn't rule out either of them as being the way the world is. Any of these scenarios regarding the compositionality of water could turn out to be the case given her epistemic situation. Importantly, Chalmers holds that this is not in conflict with the Kripkean discovery of the necessary a posteriori, since necessary a posteriori claims are claims *not* about how the world *could turn out to be*, but about how the world *could have been*. Intuitions about how the world *could turn out to be* and intuitions about how the world *could have been* come apart for the 2D theorist because they are tracking two different intensions of our token utterances. "Water could turn out to be XYZ" seems true to us because our intuitions about what utterances could turn out to be true track the primary intensions of these utterances, while "water could have been XYZ" seems false to us because our intuitions about what utterances could have been true track the secondary intensions of these utterances. Primary and secondary intensions, for the 2D theorist, are distinct functions to distinct sets of possible worlds. Primary intensions are functions from epistemic scenarios to sets of worlds, while secondary intensions are functions from worlds to sets of worlds.

Two-dimensional semantics are employed to deal with issues concerning necessary a posteriori truths, contingent a priori truths, indexicals, Kripkean puzzles about belief and so forth. For instance, take the sentence "The Métre des Archives is one meter in length", where *Métre des Archives* is a name for the platinum meter stick in Paris that for a period of time set the standard of lengths for the metric system. We'll suppose for the sake of the example that this is a bona fide example of the contingent a priori. It's true that this meter stick is one meter in length, and there is a sense in which it had to be so: after all, a meter

was stipulated to be the length of that very stick. No matter what length in inches, say, the meter stick happened to be, the causal baptism of "meter" as "the length of that stick" guarantees that that stick has to be a meter in length. But there is another sense in which the meter stick didn't have to be a meter in length. The stick could have been under immense pressure at the time of the baptism, in which case the stick would be shorter than a meter.

A two-dimensionalist tries to salvage both intuitions by holding that the sentence "The Métre des Archives is one meter in length" has both a primary and a secondary intension relative to which the truth of the sentence can be evaluated, and that the sentence comes out true on the primary way of evaluation while false on the second way. Suppose we don't know whether the meter stick was under immense pressure or not. Then, as far as we know, the meter stick could be one of any number of lengths expressed in inches. Either length in inches would be consistent with our epistemic situation. But we do know a priori that it's a meter for the reasons discussed above, irrespective of whether it was under immense pressure at the time of the baptism. We can illustrate this with the help of the following chart:

Primary↓/Secondary→	normal pressure	immense pressure
normal pressure	T	F
immense pressure	F	T

This chart gives us the *two-dimensional intension* of "The meter stick is one meter in length". A two-dimensional intension is a function from {scenario, world} pairs to truth values. Take any of these pairs and the two-dimensional intension will give you a truth value. The chart indicates that, on any scenario, the sentence comes out as actually true

but only contingently so, since the stick would be some other length than a meter had the pressure been different. The reason for this is that Métre des Archives rigidly designates *the stick*, while the term "meter" rigidly designates the *actual length* of the stick. In any world that would be counterfactual for a scenario had it turned out that that scenario were actual, the same stick would have a different length than the length denoted by "meter". This explains the contingency of the sentence.

Conceivability and metaphysical possibility

The point of discussing Chalmers' 2D framework is to investigate whether it offers the irreduciblist resources to avoid Brute Entailment. So in this section, I'll give the 2D characterization of metaphysical necessity and then discuss upshots for the modal status of the reasons fundamentalist's supervenience claim.

A sentence S is metaphysically necessary in a world w iff the secondary intension of S evaluated at w is true at all worlds. Alternatively, we can also say that a sentence S is metaphysically necessary in a world w iff S evaluated at w is true at all worlds considered as counterfactual at w.

We can know whether the secondary intension of a sentence is true at all worlds in two ways. We can know this a posteriori, or we can know this a priori. The a posteriori approach to demonstrating metaphysical necessity in Chalmers' 2D framework is just the Kripke-style causal regulation approach to demonstrating metaphysical necessity. I treat this approach separately in §3.3.2. So, in this section, I will focus exclusively on the a priori approach to demonstrating metaphysical necessity in the 2D framework.

On Chalmers' 2D conceivability framework, we can know a priori that S, given an epistemic scenario, if S is true as a matter metaphysical necessity just by knowing whether every *suitably conceivable* world that's consistent with the epistemic scenario verifies S. This is because, on Chalmers' framework, suitably conceivable worlds are metaphysically possible; therefore, if every suitably conceivable world that's consistent with an epistemic scenario verifies S, then it will be metaphysically necessary that S is true given that epistemic scenario. For example, if our epistemic scenario entails that John is a bachelor, then it is a priori that the sentence "given that John is a bachelor, John is unmarried" is true as a matter of metaphysical necessity. This is because every suitably conceivable world that's consistent with John's being an bachelor will verify John's being unmarried. Since there is no suitably conceivable world that's consistent with John's being a bachelor but does not verify John's being unmarried, it's not metaphysically possible for John to be a bachelor and also be unmarried. So it is therefore metaphysically necessary that, if John is a bachelor, then he is unmarried.

Given this framework, whether we know a priori that some S, given an epistemic scenario, is true as a matter of metaphysical necessity will depend on our criteria for determining which scenarios are suitably conceivable and therefore whether the worlds they represent are metaphysically possible. In the case of a priori knowledge, the relevant kind of conceivability will certainly be primary conceivability, or conceivability as actual. As Chalmers says, "If any variety of a priori conceivability entails possibility, it must be a variety of ideal primary conceivability" (Chalmers 2002a, p. 171). Furthermore, Chalmers asserts, plausibly enough, that "prima facie conceivability is an imperfect guide to possibility" (ibid., p. 159). For example, the scenario described in the Grim Reaper paradox

is prima facie positively conceivable. Here is the scenario: There are countably infinitely many grim reapers. Grim reaper 1 will kill you with a scythe at 1pm if and only if you are still alive at 1pm. Grim reaper 2 will kill you with a scythe at 12:30pm if any only if you are still alive at 12:30pm. Grim reaper 3 will kill you with a scythe at 12:15 pm if and only if you are still alive at 12:15pm. And so on. It's prima facie conceivable that there could be a countably infinite number of these grim reapers poised to kill you. But it's not ideally conceivable: ideal rational reflection shows that, although no grim reaper n can ever get me (because I would be killed by grim reaper $n + 1$, there must be some grim reaper that gets me at some point because I cannot survive past 12pm. This is just one example of why prima facie positive conceivability is a poor guide to metaphysical possibility.

A further question is: what kind of ideal primary conceivability provides the best guide to possibility? We have two options: *negative* ideal primary conceivability, and *positive* ideal primary conceivability. A world is negatively ideally primarily conceivable given an epistemic scenario just if the world verifies that epistemic scenario and ideal rational reflection reveals no contradiction in the world. A world is positively ideally primarily conceivable given an epistemic scenario just if it's negatively ideally primarily possible *and* it can also be coherently modally imagined, either in the way we can imagine flying horses or in the way we can imagine dark matter and ectoplasm. Alternatively, an ideally positively conceivable world just is a prima facie positively conceivable world that's also such that ideal rational reflection would reveal no contradiction in the world.

Negative and positive ideal primary conceivability can be used to construct the following two conceivability-possibility entailment theses:

> *Positive Possibility*: Positive ideal primary conceivability entails metaphysical possibility; and
>
> *Negative Possibility*: Negative ideal primary conceivability entails metaphysical possibility.

Chalmers states that Positive Possibility is "almost certainly true" and that Negative Possibility is "very likely true" (ibid., p. 172).[3]

With this background of the last three subsections in place, we are now in a position to evaluate whether the 2D characterization of metaphysical necessity provides the reasons fundamentalist with resources to avoid Brute Entailment.

Conceivability, possibility, and reasons fundamentalism

The argumentative strategy

Here again is our supervenience claim:

> *Additive Reasons Supervenience*: For any possible worlds w_1 and w_2 and any fact-agent-action-circumstance tuples T_1 in w_1 and T_2 in w_2, if T_1 and T_2 are intrinsically and extrinsically non-normatively indiscernible and w_1 and w_2 are non-normatively indiscernible, then T_1 instantiates the favoring relation if and only if T_2 instantiates the favoring relation.

Suppose Additive Reasons Supervenience expresses a metaphysical necessity in the sense that the worlds that it quantifies over are all the metaphysically possible worlds. One principle of logical entailment is universal instantiation: that, for any variable bound by a universal quantifier, you can eliminate the quantifier by substituting in a constant for the bound

[3] "Positive Possibility" and "Negative Possibility" are my terms, not Chalmers', but they do express the entailment theses that Chalmers gives at (Chalmers 2002a, p. 172).

variable. If we construct a theorem by substituting in two particular fact-agent-action-circumstance tuples (i.e. reasons-apt tuples) for the variables T_1 and T_2, this theorem alters nothing about the scope of the quantifiers binding the world variables and therefore would also express a metaphysical necessity. What this means is that, if we can show that, given the conceivability framework, one of the theorems generated from substitution does not express a metaphysical necessity, we thereby show that Additive Reasons Supervenience does not express a metaphysical necessity. This opens up a door to counterexamples to the metaphysical necessity of Additive Reasons Supervenience.

Additive supervenience relations in general state that, if two individuals are relevantly non-normatively similar, then they will be normatively similar. Additive Reasons Supervenience specifies a maximally restrictive notion of relevant similarity. In order to qualify as normatively similar on Additive Reasons Supervenience, reasons-apt tuples must be non-normatively indiscernible in every way: intrinsically non-normatively indiscernible, extrinsically non-normatively indiscernible, and in non-normatively indiscernible worlds. Call two reasons-apt tuples that are indiscernible in this way *maximally non-normatively indiscernible*, and let the *"maximal non-normative properties"* of T_1 and T_2 express the set of non-normative properties in virtue of which two reasons-apt tuples T_1 and T_2 are maximally non-normatively indiscernible.[4]

[4]This set of properties is maximally non-normative because we built Additive Reasons Supervenience specifically in order to guarantee that there were no additional descriptive properties in the universe that could function as enablers, disablers, attenuators or intensifiers for one or more of the tuples in question. If there were such additional non-normative properties, they could function in principle as enablers, disablers, attenuators and intensifiers for one of the relevantly non-normatively similar tuples but not as an enabler, disabler, attenuator or intensifier as another relevantly non-normatively similar tuple, which would then would allow us to produce counterexamples to our supervenience claim. We wanted to rule out these counterexamples in an attempt to construct the most ecumenical and easily defended additive supervenience claim that we could on behalf of the reasons fundamentalist.

Additive Reasons Supervenience will express a metaphysical necessity just if, for every instance of reason-apt tuple T at any world w_1, either every suitably conceivable world w_2 consistent with the maximally non-normative properties of T verifies T's being a reason, or every suitably conceivable world w_2 consistent with the maximally non-normative properties of T verifies T's not being a reason. In other words, the epistemic scenario relevant to evaluating the modal status of some instance of Additive Reasons Supervenience will be the epistemic scenario consisting of the maximally non-normative properties of that instance. Since the maximally non-normative properties of any instance are stipulated to include all the non-normative properties in the world containing that instance, our epistemic scenario includes knowledge of all the non-normative facts constituting the world containing the instance. If, given reasons fundamentalism, we can't suitably conceive of any two maximally non-normatively indiscernible reasons-apt tuples that differ in that one instantiates the favoring relation and the other does not, then we have found an explanation on behalf of the reasons fundamentalist of why Additive Reasons Supervenience expresses a metaphysical necessity. That explanation would simply be that we can't suitably conceive of any alternative to T's having the normative properties it does in non-normatively indiscernible worlds. The reasons fundamentalist could therefore avoid Brute Entailment.

Here is moral philosophy's favorite example of a reason-apt tuple: the fact of the drowning baby, the action of pulling the baby out of the water, the agent Peter Singer, and the circumstances in which Singer's next to the baby, the baby is the only person in immediate danger of death in the nearby vicinity, Singer is not traveling to be at the bedside of a dying family member, etc. This tuple happens to instantiate the reason relation in our world. If we substitute this reasons-apt tuple in for T_1 in Additive Reasons Supervenience, we get

the claim that any fact in some world w that there's a drowning baby just like that one, gives anyone just like Peter Singer in that world a reason to pull the baby out of the water so long as they're right next to the water, nobody else in the vicinity is in immediate danger of death, and, by the way, everything else non-normatively salient about w is also just like it is in our world in all non-normative respects.

Here is a counterexample to this Peter Singer instance of Additive Reasons Supervenience. Let w be a world non-normatively just like ours. Jamie Dimon's counterpart gets $20 million in compensation for the 2013 fiscal year for managing to limit to a mere $19.6 billion JP Morgan Chase's settlements with the SEC for the bank's role in producing the 2008 financial crisis (CNN 2014); Maximilian Kolbe's counterpart offers his life for a condemned stranger in Auschwitz; Singer's counterpart stands next to the drowning baby, and so on. The twist is that, in w, the actions Jamie Dimon's counterpart took to earn that USD$20 million are not morally dubious. Nor in w is Maximilian Kolbe's counterpart morally good, and Singer's counterpart doesn't have a reason to save the baby. In fact, nothing is morally dubious or morally good or a reason in w at all: the moral statements of its denizens are all false, and Parfit's counterpart has, by his own lights, wasted many years of his life in pursuit of ethical knowledge (Parfit 2011, section on moral naturalism). W is the *nihilistic universe*. As it does not verify any normative facts (since no normative facts exist in it), it does not verify the same set of normative facts as our universe. If the nihilistic universe is suitably conceivable, then, on Chalmers' conceivability framework, it will not be a priori that all non-normatively indiscernible possible worlds are also normatively indiscernible.

I will consider two approaches that one could take towards characterizing suitable con-

ceivability on the Chalmers conceivability framework: the Negative Possibility and Positive Possibility routes discussed earlier.

Negative Possibility

If reasons fundamentalism is true, the nihilistic universe is certainly ideally negatively primarily conceivable. This is because there is nothing about considering the nihilistic universe as actual, given the epistemic scenario consisting of all the non-normative facts of our world, that generates a contradiction, even for an ideal rational subject. After all, the conjunction of the non-normative facts of our world *and the normative facts* in our world does not generate a contradiction for an ideal thinker when considering our world as actual, so, a forteriori, the conjunction of all the non-normative facts simpliciter will not generate a contradiction for an ideal thinker either.

If there were a conceptual entailment from the non-normative features of the world to the normative features of the world, then an ideal thinker's imagining the non-normative features of the world would give rise to thoughts about the normative features of the world for that ideal thinker as a matter of conceptual necessity. For instance, if our concepts of pleasure and pain entailed that actions that increased pleasure and reduced pain were morally right, then any ideal thinker's thinking of all the pain, pleasure, and actions that produce the pains and pleasures would also produce in that ideal thinker the entailed thoughts about moral rightness.[5] However, the Primitivity component of reasons fundamentalism assures

[5]This would not be true for any thinker whatsoever. It's possible for me to think about green, for instance, without also thinking that green is a color, even though it is a conceptual truth that green is a color. It's also a conceptual truth that 29 is the tenth prime number, but I can think about 29 (such as in the sentence "I am 29 years old") without thinking that 29 is the tenth prime number. In fact, I think I have never thought about 29's being the tenth prime number before today, despite the fact of my having been 29 for a while now. However, we aren't talking about me here. We're taking about an ideal thinker. Ideal thinkers are stipulated

that there is no conceptual entailment from non-normative concepts to normative concepts. If there were such a conceptual entailment, then we would have at least a partial analysis of our concept of the favoring relation in terms of the non-normative concepts that would conceptually entail thoughts about the normative concepts by an ideal thinker.

Chalmers claims that Negative Possibility—in other words, the claim that ideal primary negative conceivability entails primary possibility—is "very likely true". If Negative Possibility is true, then the nihilistic universe is suitably conceivable and we have a counterexample to the metaphysical necessity of Additive Reasons Supervenience.

However, Chalmers explicitly cites the irreduciblist primitivist conception of normative concepts as a possible counterexample to his Negative Possibility thesis. He states that "there do not seem to be distinct positively conceivable situations" in which both situations verify the non-normative truths but only one verifies the normative truths. If, as reasons fundamentalists think, the moral truths were not determined a priori by the non-normative truths, we would then have a case in which the nihilistic universe is negatively ideally primarily conceivable but not possible, which would constitute an objection to Negative Possibility. This is a problem for Chalmers, since Chalmers wants Negative Possibility to be true.

Chalmers himself does not have much to say regarding moral truths. (I was once in a seminar of Derek Parfit's in which Parfit asked Chalmers, "What do you think?" Chalmers replied, "About what? Normativity? Too hard for me!") But he does say a few things. In particular, Chalmers suggests that the moral truths do not pose as much of a problem for reduciblism about normativity as the phenomenal truths pose for reductivism in the

to be thinkers that, unlike me, always think through all the conceptual implications of their thoughts.

philosophy of mind:

> "There are two disanalogies [between the moral and phenomenal truths], however. First, there does not seem to be a conceivable world that is naturally identical to ours but morally distinct, so it is unlikely that moral facts are further facts in any strong sense. Second, moral facts are not phenomena that force themselves on us. When it comes to the crunch, we can deny that moral facts exist at all... The same strategy cannot be taken for phenomenal properties, whose existence is forced upon us" (Chalmers. 1996, p. 83).

Chalmers here notes the potential tension between the Negative Possibility thesis and the irreduciblist primitivist position on normative truths. He does not much care about normativity, so he is happy to either reduce it (such that the normative facts are not "further facts in any strong sense") or eliminate it altogether. The irreduciblist obviously can't reduce normativity or eliminate normativity. So the irreduciblist would have to reject the Negative Possibility thesis and assert only the Positive Possibility thesis: that, for any scenario, only if that scenario is positively ideally primarily conceivable can we conclude that worlds consistent with that scenario are metaphysically possible. This brings us to an examination of Positive Possibility and its implications for the metaphysical necessity of Additive Reasons Supervenience.

Positive Possibility

I mentioned in §3.3.1 above that, for Chalmers, a scenario is positively ideally conceivable if (a) we can picture the world in our mind's eye like we can picture flying horses, or (b) we can't picture it in our mind's eye, but nevertheless positively imagine it in a way similar to the way we imagine ectoplasm[6] or dark matter. The favoring relation is not something that

[6]Ectoplasm is a theoretical posit in the philosophy of mind literature. It is supposed to be some immaterial substance that instantiates phenomenal properties and is frequently used in thought experiments about the

we can visualize in either way. On the reasons fundamentalist picture, the favoring relation does not have empirical properties, so it's not something that we can visualize. And it isn't located in space like ectoplasm or dark matter, so it's not something we can conceive of in that way either.

However, Chalmers does not actually think that a scenario is positively ideally conceivable only if you can imagine it in at least one of these two ways. Chalmers also thinks that we can positively imagine mathematical truths, and the facts that these truths denote also aren't located in space or have empirical properties. Although the criteria for positive conceivability are not well worked-out, we should assume that, whatever these criteria end up turning out to be, we should be able to positively imagine at least actual normative scenarios, such as the scenario in which the fact of the drowning baby is a reason for Singer to save it. The question, then, is whether the nihilistic universe construal of the Peter Singer scenario—namely, the scenario in which fact of the drowning baby is *not* a reason for Singer to save it—is positively conceivable in the same way.

I don't actually know the answer to this. However, I think there is a good case to be made that, if reasons fundamentalism is true, the nihilistic construal of the Singer scenario is positively ideally conceivable. If it is positively ideally conceivable, then Positive Possibility does not give reasons fundamentalists resources to avoid Brute Entailment.

I will give two arguments for the claim that Positive Possibility does not give reasons fundamentalists resources to avoid Brute Entailment. The first argument will adapt Chalmers' arguments for the positive conceivability of phenomenal zombies to the case of normativity. The second will adapt Michael Smith's permutation problem for Jackson's an-

modal character of phenomenal properties.

alytic moral functionalism to the case of reasons fundamentalism. As I said, I take neither argument to be conclusive.

Phenomenal zombies . A phenomenal zombie, for Chalmers,

> is just something physically identical to me, but which has no conscious experience—all is dark inside. While this is probably empirically impossible, it certainly seems that a coherent situation is described; I can discern no contradiction in the description.... Almost everybody, it seems to me, is capable of conceiving of this possibility. (ibid., p. 96).

Chalmers furthermore suggests that phenomenal zombies are positively conceivable. In other words, for Chalmers, we not only can imagine a world featuring his zombie twin, but also that the zombie twin is positively absent of phenomenal experience. There is not merely an *absence of imagining* the phenomenal properties, but an *imagination of the absence* of these properties. Thus it is possible on the Chalmers conceivability framework for absences to be positively conceivable.

In his 1996 book, Chalmers gives a sustained defense of the positive conceivability of phenomenal zombies as part of his overall argument against reductive views in the philosophy of mind. On one such reductive view, which Chalmers in later work calls "Type-C" materialism[7] (Chalmers 2002b), there appears to be a deep epistemic gap between the physical and phenomenal domains, but this gap is closeable in principle. Although it would have been inconceivable to the pre-Socratics, for example, how matter could be energy, we later experienced a conceptual revolution that allowed us to conceive this very thing. Similarly, on type-C materialism, it is currently inconceivable to us how consciousness could be phys-

[7]In this discussion about reductivist views in the philosophy of mind, I do not distinguish between "materialism", "physicalism", and "reductivism". The terms are used synonymously.

ical, but there will in time come a conceptual revolution that will allow us to understand this very thing. That's the type-C reductivist view.

Another way to gloss the type-C reductivism view is as the claim that, although phenomenal zombies are prima facie positively conceivable, they are not ideally conceivable. They are not ideally conceivable according to type-C reductivism since, if we were ideal thinkers, we would have access to the conceptual capacities that would allow us to see how consciousness could be physical, in which case we would not be able to conceive of a physical duplicate of myself that lacked consciousness.

Chalmers argues that the argument against the ideal positive conceivability of phenomenal zombies is unsound. His reasoning is simply that physical explanation is always given in terms of structure and dynamics, and that there is nothing about structure and dynamics that could explain phenomenal properties. No matter what the future physics looked like, it would at least be a theory that purported to explain physical reality in terms of structure and dynamics. Since these kinds of concepts are not suited in principle to explain the existence of phenomenal properties, we should accept that there is would be no contradiction that such an ideal thinker could generate if she were to try to conceive of phenomenal zombies. Phenomenal zombies are therefore positively ideally conceivable.

A similar worry appears in the normative case. Like phenomenal zombies, it seems that we can prima facie positively imagine the absence of normativity. Consider this interchange that Primo Levi describes between himself and a guard at the camp in Auschwitz:

> Driven by thirst, I eyed a fine icicle outside the window, within hand's reach. I opened the window and broke off the icicle but at once a large, heavy guard prowling outside brutally snatched it away from me. "Warum?" I asked him

> in my poor German. "Hier ist kein warum" (there is no why here), he replied, pushing me inside with a shove. (Levi 1993, p. 26)

In asking "warum?", Levi is asking the guard for a practical normative reason, some sort of justification for the guard's behavior. The guard's response was not "I didn't do it for any reason at all", which would merely indicate something about the current absence of motivating reasons on the part of the guard. His response to Levi, rather, is simply, "there is no why here."

I suspect that most of us know what that guard means. He means that there are no reasons. He understands what Levi is asking him: that Levi's request is a request to be given a normative reason for the guard's behavior, and that Levi is ultimately looking for some sort of normative framework from which to understand the normatively incomprehensible world in which he finds himself. Thus the guard says not simply "there is no why", which would indicate simply that there were no normative reason for ripping the icicle from Levi. The guard says, rather, that there is no why "here": not just in this minor act of cruelty, but in the entire camp. The entire camp is devoid of normativity. There are orders and threats, action and submission. But there are no reasons. Not here.[8]

We can't understand the absence of normativity in this scenario simply as the guard's

[8]This dovetails with a claim that Parfit made in seminar. Parfit indicated that the discovery of reasons is particularly important because the kind of practical reasoning that characterizes people who lack this concept can sometimes manifest as reasoning from duties and commands. This kind of moral reasoning, Parfit suspects, was the primary mode of moral reasoning of soldiers and government officials serving in the Third Reich. The theory of reasons is fundamentally opposed to a theory of duties and commands. You can, of course, have reasons to obey commands, but you don't need reasons to obey a command. This is because commands are imperatives that issue from some authority that stands outside of your rational agency. (Ignore Kant for our purposes here. We are not talking about moral rational imperatives but about commands issued by authoritarian figures in positions of political power.) Acting from reasons, in contrast, always issues from your own rational point of view and as such is not an entirely alienating experience. The discovery of reasons, therefore, may help society mitigate the frequency of the kind of alienated moral experience that gave rise to the horrors of the Nazi regime and could give rise to future authoritarianism as well.

having an absence of normative imagination. The fact that the guard understood what Levi is asking disallows this. In understanding what Levi was asking but stating that there was no why in Auschwitz, the guard is expressing not that he is failing to imagine anything normative at all, but that he is positively imagining the normativity's absence. The case, therefore, seems to suggest the prima facie positive conceivability of the nihilistic universe, at least on a small scale.

A "normative type-C reductivist" might say that, although there is currently a conceptual gap between the non-normative properties and the normative properties, future empirical research could initiate a conceptual revolution in our non-normative concepts that would allow us to see how the non-normative features of the world give rise to the normative features that they do. If this were the case, then the nihilistic universe, for example, would be prima facie positively conceivable but not ideally conceivable, since the ideal thinker would have access to the concepts available after the conceptual revolution and would be able to use these concepts to understand how the non-normative features of the world give rise to the normative features that they do.

However, the non-normative features of the world, for the reasons fundamentalist, are different than the normative features of the world in an important and radical way: they are *non-normative*. Non-normative explanation is in principle unsuited to explain the existence of normativity. There can be no conceptual revolution that could give us concepts to bridge this explanatory gap. For this reason, there is nothing about the prima facie positive conceivability of the nihilistic universe that prevents an ideal thinker from conceiving it as well, in which case the nihilistic universe is not only prima facie positively conceivable but ideally conceivable as well.

The nihilistic universe was stipulated to be indiscernible from our world in all non-normative respects. If both the nihilistic universe and our own world, with its own normative features, are ideally positively conceivable, then, given the Positive Possibility thesis that positive ideal conceivability entails metaphysical possibility, we have two metaphysically possible worlds that are indiscernible with respect to their non-normative features but discernible with respect to their normative features. Additive Reasons Supervenience entails that no two non-normatively indiscernible worlds can differ with respect to their normative features. Therefore, on the conceivability framework, Additive Reasons Supervenience cannot express a metaphysical necessity, and Brute Entailment remains a problem for reasons fundamentalism.

I'll now move on to the second argument for why Positive Possibility does not help the reasons fundamentalist avoid Brute Entailment.

The permutation problem . Reasons fundamentalists hold that the reasons relation is primitive. This is to say that, according to reasons fundamentalism, is no direct conceptual entailment from our concept of a reason to any conclusion about whether any particular reasons-apt tuple instantiates the reasons relation. If the reasons fundamentalist wants to explain why some reasons-apt tuple is a reason, he needs to appeal to something in addition to the concept of a reason.

Scanlon's own particular approach to explaining why particular reasons-apt tuples are reasons appeals to what he calls the *pure normative truths* (Scanlon 2014, pp. 37-42). Pure normative truths are metaphysically necessary truths that, along with the help of "mixed" normative-nonnormative bridge laws, allow us to explain why any particular reasons-apt

tuple is a reason. One example of a pure normative truth is "pain is bad". When I rationally judge that I have a reason not to slap Johnnie, what I am doing, according to Scanlon, is grasping the pure normative truth that pain is bad, and then inferring from this and a mixed normative truth such as "slapping people causes pain" to the conclusion that, for example, I have a reason not to slap Johnnie even though I am really mad at him right now for not doing the dishes.[9]

Pure normative truths do not admit of any explanation of their truth. Pain is bad; that's all there is to it. This is problematic for our purposes, because what we are trying to do in this section is to find a way to allow the reasons fundamentalist to avoid Brute Entailment. If we do not have any way to explain why "pain is bad" is true, we also don't have a way to explain why "pain is bad" is true as a matter of metaphysical necessity. Since all explanations of why particular reasons-apt tuples are reasons involve an appeal to the pure normative truths, the result is that we also lack an explanation of why any particular reasons-apt tuple's being a reason obtains with metaphysical necessity.

The pure normative truth approach is not the only approach that the reasons fundamentalist can take. It is open to the reasons fundamentalist to be an explanatory coherentist instead. The reasons fundamentalist can say that we can explain why any particular reasons-apt tuple is a reason by appealing to some other reasons fact along with additional non-normative facts. We can explain why Johnnie's skipping his dishes duty is not a reason to slap Johnnie by appealing to some other reason, like "skipping dishes is not a reason to cause people pain" and the non-normative fact that stabbing causes pain. We can explain

[9]Scanlon claims that we don't have to consciously grasp the pure normative truth. He takes his view to be consistent with our common experience of just "seeing" the reason and acting (or refraining from acting) immediately, without having done any further conscious practical reasoning beforehand.

why this is a reason by citing the fact that "infrequent forgetfulness don't provide reasons to cause people pain" and the non-normative fact that Johnnie forgot this time but he usually remembers. We can explain why this is a reason by saying "infrequent forgetfulness doesn't provide reasons to cause roommates pain" and the non-normative fact that Johnnie is a roommate. And we can explain this with "take a petty action like slapping my roommate Johnnie for forgetting dishes duty—I definitely don't have a reason to do that!" (Why not? Because it will cause Johnnie pain, and so on.) In reality, the explanation scheme for why any particular reasons-apt tuple is a reason will be more involved than this, but this example at least illustrates what the explanatory coherentist scheme I'm proposing would look like.

Here is the problem, which I adapt from a similar argument given against moral functionalism by Michael Smith at (Smith 1994, pp. 48-55). You can take this entire explanatory structure, but permute it into an explanation of why Johnnie's skipping his dishes duty *is* a reason for me to slap Johnnie. Here is what the permuted structure would look like. Jones claims that Johnnie's skipping his dishes duty *is* a reason for me to slap Johnnie because slapping Johnnie will cause Johnnie pain, and Johnnie's skipping his dishes duty is a reason to cause Johnnie pain. This is a reason, Jones claims, because Johnnie infrequently forgets, and this infrequent forgetfulness is a reason to cause Johnnie pain. Why? Because Johnnie is my roommate, and if your roommate is infrequently forgetful, you should cause them pain. Why? Well, Jones claims, my roommate Johnnie once forgot to do the dishes, and that gave me a reason to slap him. Call the world as Jones sees it the *permuted world*.

The general idea behind permutation is this. Let the *reasons valence* of a reasons-apt tuple be the reasons-apt tuple's status as a reason. If a reasons-apt tuple has a positive

reasons valence, it is a reason; if it has a negative reasons valence; it is not a reason. On the coherentist scheme, the reasons valence of any particular reasons-apt tuple can be explained by appealing to the valence of some other reasons-apt tuple, along with, in some cases, some further non-normative facts. What I am suggesting is that there may be, in principle, no (theoretical) reason why the reasons valences of an entire coherentist explanatory scheme can't be permuted in the way that I have just illustrated.

I'll explain this with the help of a chart featuring all the reasons-apt tuples and their valences from our Johnnie example. I leave out the agents and circumstances because they don't vary in this case, and I include the non-normative facts that are relevant to the explanations.

Fact	R-Valence	Action	Non-normative
J skips dishes	No	slap J	N/A
J skips dishes	No	cause J pain	stabbing $\rightarrow$ pain
forgetting	No	cause J pain	J forgot
forgetting	No	cause pain to roommate	J is a roommate
J skips dishes	No	slap J	My roommate J forgot to do dishes

Here is what the chart looks like after we permute all the reasons valences:

Fact	R-Valence	Action	Non-normative
J skips dishes	Yes	slap J	N/A
J skips dishes	Yes	cause J pain	stabbing $\rightarrow$ pain
forgetting	Yes	cause J pain	J forgot
forgetting	Yes	cause pain to roommates	J is a roommate
J skips dishes	Yes	slap J	My roommate J forgot to do dishes

As far as I can tell, both charts depict a coherent normative explanatory scheme. The valence of each reason in each scheme has an explanation that appeals to the valence of another reason in the scheme. The same non-normative facts appear in both explanatory schemes, and this set of non-normative facts is mutually consistent. Other than the reasons valences, there is no normative content in either explanatory scheme that one could use to demonstrate some sort of inconsistency within the normative claims. On the reasons fundamentalist picture, furthermore, the concept of a reason is primitive. In itself, it gives us no conceptual resources that we could use to show that the normative facts that exist in the permuted world are inconsistent with the non-normative facts of that world. In order to show such inconsistencies, one would have to find these inconsistencies from within the normative explanatory scheme itself. However, there do not seem to be any such inconsistencies. There is no clear way how you could convince someone who held these normative beliefs why, by his own lights, he does not have an internally consistent and coherent theory of which reasons-apt tuples are reasons and which are not.

What the possibility of permuting the reasons valences in an entire normative explanatory scheme suggests is that there are distinct, internally coherent theories of which reasons-

apt tuples are reasons and which are not that generate distinct answers to questions about whether any particular reasons-apt tuple is a reason or not. If these distinct explanatory schemes are ideally positively conceivable, then, given the thesis that positive ideal conceivability entails metaphysical possibility, we have a counterexample to Additive Reasons Supervenience.

Furthermore, we do have some reason to think that both of the sample explanatory schemes that I have are ideally positively conceivable. On Chalmers' picture, an ideally positively conceivable scenario is just a prima facie positively conceivable scenario that reveals no contradiction even after rational reflection. Both of the explanatory schemes I've given are prima facie positively conceivable. If Jones told us that the permuted world was actually our own, we would think that Jones is certainly incorrect, not a very nice person, and possibly mentally ill. But we would understand what he said, and we can even get into the mindset of what it might be like for someone to conceive of normative reality as he does. Given the non-normative facts of our world, his view of normative reality would strike us as false but positively conceivable, at least prima facie. Furthermore, if the permuted world is prima facie positively conceivable, and the permuted world's normative explanatory scheme was internally coherent, the permuted world must be ideally positively conceivable as well. For without any internal inconsistency in the normative theory that describes the normative reality of the permuted world, there would be no internal contradictions that an ideal thinker could detect.

The explanatory scheme of the permuted world and the explanatory scheme that correctly expresses the normative nature of our world are mutually inconsistent. For example, in the permuted world, your roommate's failure to do dishes is a reason to slap him, whereas

in our world, your roommate's failure to do dishes is not a reason to slap him. It can't be the case that both are true. Because of this, it can't be the case that there is a single possible world that verifies both of them. If we can positively ideally conceive of either scheme, we must be ideally positively conceiving of distinct worlds, in which case the conceivability approach to metaphysical necessity would entail that there are two metaphysically possible worlds which are non-normatively indiscernible but discernible with respect to the distribution of the favoring relation in each world. If this were the case, Additive Reasons Supervenience, which entails that two non-normatively indiscernible worlds cannot be discernible with respect to the distribution of the favoring relation, could not express a metaphysical necessity. This means that the conceivability approach to metaphysical necessity does not help the reasons fundamentalist avoid Brute Entailment.

Some caveats. I have claimed that the permuted world is prima facie positively conceivable. However, I am not claiming that the permuted world is ideally positively conceivable. I am only claiming that there do not seem to be resources available to the reasons fundamentalist to resist the suggestion that the permuted world is not ideally positively conceivable. If a different theory of normativity N_T were true, it could well be the case that, despite the permuted world's prima facie positive conceivability, the permuted world would not be ideally conceivable for an ideal thinker who understood the true theory N_T in virtue of her ideal thinkerhood.

Secondly, I don't know if the permutation scheme would work for an entire coherentist explanatory scheme. More work needs to be done on that question. There is indeed a debate on whether there is a permutation problem for Jackson's network analysis of normativity, and my sense from this debate is that there are enough conceptual resources in Jackson's

theory to respond to this problem.[10] However, no one, as far as I know, has attempted to investigate the existence of this related permutation problem for the reasons fundamentalist. Furthermore, the permutation problem for Jackson's network analysis of normativity is supposed to be a problem for Jackson's reduciblism, while my rendition of the problem for the reasons fundamentalist is meant to describe a problem for the reasons fundamentalist's use of the conceivability approach to metaphysical possibility to avoid Brute Entailment. I do, however, take the prospects for developing my permutation problem for the reasons fundamentalist to be better than the prospects for developing a permutation problem for network analyses of normativity in general, since network analyses of normativity typically allow for a whole host of thick normative concepts and complex normative concepts that can be used to generate conceptual contradictions. Reasons fundamentalism, however, has no conceptual resources to resist this outside of the single primitive concept of the favoring relation.

Another problem with my argument here is the possibility of normative insularity. Rovane argues that relativists about normativity are committed to the existence of distinct, normatively insulated bodies of normative truths (Rovane 2013, p. 79). Sets of normative truths Γ and H are *normatively insulated* from each other just if (a) any pair of sentences γ_1 and γ_2 in Γ stand in logical relations with each other, (b) any pair of sentences η_1 and η_2 in H stand in logical relations with each other, and (c) there exists no pair of sentences γ in Γ and η in H such that γ and η stand in logical relations with each other. In other words, sets of normatively insulated truths form equivalence classes in which the equivalence relation partitioning the truths into the equivalence classes is the relation of licensing logical infer-

[10]For more on this, see, for example, (Lenman 2006, §4.3).

ence. For example, any two sentences γ_1 and γ_2 in Γ will license the following inference: $\gamma_1 \vdash \gamma_1 \vee \gamma_2$. However, if Γ and H are normatively insulated, there is no pair of sentences γ in Γ and η in H that licenses $\gamma \vdash \gamma \vee \eta$.

Normative insularity is a problem for my claim that the ideal positive conceivability of the permuted world because it could be the case that the bodies of normative truths that feature in the two coherent explanatory schemes above are mutually normatively insulated bodies of truths in our world. In order words, although it's true that Johnnie's not doing the dishes is *not* a reason to slap him for all the reasons I have given, it might also be the case that his not doing the dishes *is* a reason to slap him for all the reasons I've given, but that these claims about reasons are normatively insulated and therefore do not license the kinds of inferences that would show them to be mutually inconsistent. If we can positively ideally imagine their normative insularity, then we no longer have to conclude that, when we imagine each scenario expressed in the two coherent explanatory schemes, we are imagining distinct metaphysically possible worlds. Our ability to ideally positively conceive of the permuted world would therefore fail to generate a counterexample to the claim that Additive Reasons Supervenience expresses a metaphysical necessity.

Whether the normative insularity objection succeeds against my argument depends on (a) whether there are such things as normatively insulated bodies of truths, and on (b) whether distinct bodies of truths are simultaneously ideally positively conceivable. Whether this is the case is an area for further investigation, and one that I will not carry out in this dissertation because my dissertation is already too long as it is.[11]

[11] See (Rovane 2013) for a detailed discussion of whether there are or could be normatively insulated bodies of normative truths.

In this section §3.3.1, I have argued that the reasons fundamentalist has significant hurdles to overcome in order to use the conceivability approach to explaining metaphysical necessity as a means to avoid Brute Entailment. In the next section §3.3.2, I will examine the reasons fundamentalist's prospects for avoiding Brute Entailment with the help of a posteriori approaches to explaining metaphysical necessity.

4.2 Causal regulation

In this section, I consider the implications of a posteriori approaches to metaphysical necessity on the prospects for reasons fundamentalists to avoid Brute Entailment. The a posteriori approach is, of course, the avenue that Brink, Boyd and Copp in particular have elaborated and defended in work done through the past 25 years.

There are a number of outstanding problems with these views, ones that I believe are at this point definitive but have been sufficiently elaborated in other work.[12] The question in place for us here is not whether these views are defensible, but whether they give the reasons fundamentalist resources avoid Brute Entailment. I think they do not. To argue for this, I will consider the implications of this approach to explaining metaphysical necessity through a critical analysis of David Brink's normative irreduciblism, which I take to be representative of the a posteriori approach to explaining normative supervenience.

[12] In particular, the views have trouble accounting for the intuitive force of counterexamples and intuitions about transworld disagreement (Horgan and Timmons 1991). Ken Yasenchuk has compiled a helpful dissertation on the subject at (Yasenchuk 1995). Another helpful dissertation is (Rubin 2009).

What is the causal regulation of reference?

Some a posteriori truths are necessary truths. Truths about the nature of the referents of natural kind terms and proper names are two examples of such necessary a posteriori truths.

Putnam argued that natural kind terms acquire reference in two ways. They may acquire reference through ostension, in which the linguistic community is presented with samples of the natural kind and stipulates through a kind of speech act that the natural kind term is to apply to that and only that natural kind presented in the samples. 16th century English speakers go to the Indian subcontinent and see tigers for the first time; they declare that whatever those things are shall be called "tiger". The English term "tiger" thereby acquires its reference as the natural biological kind that was baptized in this initial ostension procedure. The reference of such terms is said to be "causally regulated", since there is a complex causal chain that connects the reference of any particular use of the term "tiger" today to that first "dubbing ceremony" when "tiger" gained its reference.

Natural kind terms may also acquire reference by being associated with a particular description and gaining their reference by being stipulated to denote the natural kind that uniquely satisfies that description. Primitive English speakers say that "gold" is whatever that yellowish stuff is that they put on all the religious artifacts; the English term "gold" thereby acquires its reference as the natural kind that happened to be on all the religious artifacts—namely, gold.

Since the ostension and description-specifying procedures fix the reference relations of terms to natural kinds, the reference relations connect these terms to the natural kinds even when these natural kinds no longer satisfy the initial reference-fixing description or the

descriptions that would have been satisfied by the original samples. "Tiger" still refers to the tiger it refers to after you paint that tiger maroon; "gold" still refers to gold even though contemporary churches frequently just use gold-colored paint they picked up on discount at Sherman Williams. Importantly, natural kind terms do this even in non-actual worlds. So when we are talking about tigers and gold in possible, non-actual worlds, our words pick out the natural kind tiger and the natural kind gold when we use these natural kind terms to discuss these non-actual worlds as well. "Gold" and "tiger" refer to gold and tigers even when we talk about worlds in which all the gold and tigers in that world are painted maroon exceptionlessly. This is to say that the reference-fixing of natural kind terms specifies the reference of natural kind terms necessarily: the terms refer to the same natural kind in all possible worlds, even the ones in which all the tigers and gold are painted maroon. This feature of natural kinds terms is called their *rigid designation*: in other words, when natural kind terms whose reference is fixed in one of the two ways specified successfully refer to a natural kind, they refer to that natural kind necessarily.

Since the ostension or description-specifying procedures fix the reference relations to natural kinds necessarily, anything we discover about these natural kinds through empirical investigation will be a discovery of necessary truths. For example, if we discover that gold consists of atoms with atomic number 79, we learned that "what it is to be gold is to consist of atoms with atomic number 79" is true. But we didn't simply learn that "gold consists of atoms with atomic number 79" is true; we learned that it is true necessarily, since "gold" refers to gold necessarily, and what it is to be gold is to consist of atoms with atomic number 79.

Ostension and definition-specifying procedures can also fix the references of proper

names, which opens the door to empirical discovery of necessary truths about individuals. Thus the ostension procedures that rigidly designated the references of "Hesperus" and "Phosphorus" both as the planet Venus opened the door for empirical discovery that Hesperus and Phosphorus were identical, and therefore, given the rigidity of "Hesperus" and "Phosphorus", necessarily identical as well.

Causal regulation of the reference of normative terms

The fact that natural kind terms and proper names are rigid designators raises the question whether other terms can rigidly designate their referents as well. Brink argues that normative terms do. He endorses, for instance, the following thesis:

> Each moral term *t* rigidly designates the [non-normative] property *N* that uniquely causally regulates the use of *t* by humans. (Brink 2001, p. 165)[13]

Brink not only has this causal regulation view of the reference of normative terms, but also has a view as to what properties causally regulate the reference of these normative terms. Take all of the properties that stand under our non-normative concepts. Call these the *descriptive properties*. Brink argues that no descriptive property or concatenation of descriptive properties can causally regulate the reference of normative terms.[14]

To defend this claim, Brink gives an argument from the "multiple realizability" of normative properties. Properties are multiply realizable just if they are realizable by not one property but multiple properties. Pain states are frequently said to be multiply realizable

[13]Brink here is actually quoting from (Horgan and Timmons 1991), but Brink also accepts this thesis.

[14]This does not mean that Brink is a non-naturalist. Brink thinks that he is ultimately a naturalist because he holds that normative properties *are* natural properties, where the "*are*" here is the "are' of constitution rather than the "are" of identity.

in this way. Pain states can be realized by C-fiber firings, but they can also be realized by firings of very different neural structures and even, most people think, activity of non-carbon-based lifeforms. Brink thinks that normative properties are multiply realizable as well. As Brink writes,

> Both the property of injustice and the particular instances of injustice, in whatever social or economic conditions they are actually realized, could have been realized by a variety of somewhat different configurations of social and economic properties and property instances. Moral properties could have been realized by an indefinite and perhaps infinite number of sets of natural [i.e. descriptive] properties. If we deny that identity is a relation that can hold between relata that are indefinitely or infinitely disjunctive—say, because we insist that identity holds only between genuine properties and we deny that disjunctive properties are genuine properties (Armstrong 1978, pp. 19-23)—then the multiple realizability of moral properties provides us with a reason for resisting the identification of moral and natural [i.e. descriptive] properties. (Brink 1989, pp. 158-159)

Here is Brink's argument put differently. Normative properties are multiply realizable. What this means is that, in our world, there will be a certain concatenation of economic and social properties that realize the property of injustice. However, there will be other worlds in which these specific economic and social properties will not realize the property of injustice, and worlds in which other economic and social properties do realize the property of injustice. For example, the economic condition in which Nestle owns all the water rights in large swaths of Kenya realizes injustice in our world, but in worlds in which people in Kenya didn't need water or had money to pay Nestle, the economic condition in which Nestle owns all the water rights in large swaths of Kenya perhaps would not realize the property of injustice. Furthermore, the economic condition of harvesting seaweed to sell to consumers does not realize the property of injustice in our world, but they might in

worlds in which seaweed were sentient. Now, we could say that the property that causally regulates "injustice" was some disjunctive property, in which case we could say that the property that causally regulates "injustice" just was the descriptive property "monopolizing of water access or harvesting sentient seaweed or ... or ..." However, there are no disjunctive properties. So we cannot say that this disjunctive descriptive property is the property that uniquely causally regulates our use of "injustice". The argument extends mutatis mutandis for the other normative properties.

One could take this argument as an argument against the causal regulation theory of the reference of normative terms. Brink, however, does not do this. Instead, he claims that the upshot of this argument is that "injustice" is causally regulated by—surprise—injustice. But injustice is not identical to a descriptive property or some concatenation of descriptive properties. It is, rather, fully constituted by but not identical to such properties (ibid., p. 158). Ditto for all the other normative properties.

Brink's view is that injustice and the other normative properties are numerically distinct from any properties that can be denoted by our non-normative concepts. Brink's view of normativity therefore is irreduciblist on the construal of irreduciblism that I gave in §2.5 above.

Brink's view would count as a *naturalism* if one understood naturalism to be, as some people do, the view that ethical properties are the object of study of the natural sciences. Brink does think that the nature and identity of the normative properties are the object of study of the natural sciences, in two ways. First, Brink states[15] that the most appropriate way to determine the identity of normative properties like injustice is to use scientific means

[15]In earlier work. His later work seems to move away from this view, for instance in (Brink 2001).

of investigation into the identity of the properties that causally regulate our use of normative terms. Secondly, once we have so identified these properties, we can use scientific means of investigation to further discover the nature of these properties. I indicated in my discussion of irreducibility, I am loathe to use the term "naturalism" because I believe it to be ambiguous in the metaethical context, and because I worry that the term conflates metaphysical and epistemological features of views that should be kept conceptually distinct. Readers that do not have these qualms, or are otherwise willing to stipulate that natural properties are properties whose identity and nature are suitable objects of scientific investigation, can safely understand Brink's view as a naturalism about normative properties.

I am neutral on whether there are disjunctive properties. However, if there were disjunctive properties, then the causal regulation view that would result from identifying the normative properties with these disjunctive properties would give us a reduciblism, not an irreduciblism, about normative properties. This chapter only concerns the prospects for explaining normative supervenience for irreduciblist theories of normativity. So we can ignore the case in which there are disjunctive properties. We can also ignore the case in which the normative properties turn out not to be multiply realizable, because this too would result in a reduciblist view given the structure of Brink's argument.

What I have just described is a formal aspect of Brink's theory of normativity. Brink's substantive theory of moral properties can be classified as a *moral role functionalism* about moral properties. For Brink, moral properties are to be identified with the higher-order properties of playing particular causal roles in specific characteristic activities of human organisms (Brink 1984, p. 119). Those specific characteristic activities are those which "bear upon the maintenance and flourishing of human organisms" (ibid., p. 120), where

flourishing is spelled out in some suitably descriptive way. For example, in our world, there is a complex causal system *S* that serves to distribute food and water to human beings. *S* will have various lower-order descriptive properties, but goodness, according to Brink, is not to be identified with any of these. Goodness is, rather, to be identified with the following higher-order property of *S*: being such that *S* bears upon the maintenance and flourishing of human organisms.

It's probably most helpful at this point to explain Brink's view through a comparison between it and the more familiar discussions of functionalism in the philosophy of mind. In philosophy of mind discussions of functionalism, there is a familiar distinction that is drawn between "role" functionalism and "realizer" functionalism. Both sorts of functionalism maintain that mental states are to be identified with properties that play a particular role in some larger functional system.[16] In the standard causal functionalist case, mental properties would be identified with properties that play a particular role in a causal system. The distinction between role and realizer functionalism has to do with how one prefers to interpret the phrase "properties that play a particular role". For the *realizer functionalist*, the mental properties are identical to the lower-order properties that happen to realize the mental property in question. For the *role functionalist*, the mental properties are identified with the higher-order properties of *being such* that they play the particular role.

Lest this be interpreted to be a Byzantine and unimportant distinction, here is an illustration. At certain points in the development of neuroscience, it was thought that pain

[16]Frequently, these systems are causal systems, but they need not be. For instance, the imaginary number in complex analysis is sometimes said to be defined functionally: as the number such that its square is -1. Similarly, the natural numbers in Peano arithmetic are defined in terms of the number 0 and applications of the successor function, and so on.

sensations were accompanied always by firings of C-fibers. Let's pretend for a moment that this is true. JJC Smart is well known in part for his suggestion that pain was not only accompanied by C-fiber firings but identical to them. Such a view would qualify as a type-identity theory of the mind. A realizer functionalist theory might instead say that pain is identical, not to the C-fiber firings right off the bat, but to whatever happens to play the pain role in the sophisticated causal structure that is disposed to produce some specified range of characteristic human behavior. The pain role might be, say, an event brought about by bodily injury and such as to produce anxiety, beliefs in one that something is wrong with one's body, the desire to be out of that state, and moaning or wincing behavior.[17] Pain would then be identical to C-fiber firings on the realizer functionalist view just if it turned out that C-fiber firings played this causal role in the sophisticated causal structure that characterizes human activity.

Block's influential criticism of realizer functionalism had to do with the theory's purported chauvinism (Block 1978). Suppose that there were an alien species composed not of carbon but of silicon, or whose nerve fibers were optical fibers as opposed to neurons. If realizer functionalism identifies various mental states, such as pain, with the firings of C-fibers, then it seems that members of the aforementioned alien species lack mental states. After all, the alien species doesn't have any C-fibers, per hypothesi, so if there's nothing more to pain states than being C-fiber firings, the alien species cannot be in pain states. Is it metaphysically impossible for aliens with a physiology quite different from us to have pain states? Most of us don't think so.

Role functionalism lacks this problem. According to role functionalism, mental prop-

[17] I draw this description of pain's causal relations from (Levin 2013).

erties are not identified with the lower-order properties that play the appropriate role in the causal functional system, but with the higher-order properties of being such that they play this role. On role functionalism, the pain of the alien species would be realized by whatever part of them happened to play the appropriate pain-causal role in the sophisticated causal structure that characterized the alien's activities. This opens up the possibility that some aspect of the alien physiology, although quite different from ours, could still play the appropriate mental causal role.

Brink's explanation of normative supervenience

Brink goes on to argue that this not-identity-but-constitution view of the relation between normative properties and descriptive properties provides a ready explanation of normative supervenience. Supervenience relations are kinds of necessitation relations: if B supervenes on A, then whenever you have A you must have B. Injustice, per hypothesi, is a multiply realizable property. This means that there are many discrete concatenations of descriptive properties that realize injustice. Call these discrete concatenations of descriptive properties that realize injustice $N_1, N_2, ...N_n$. Since each N_i realizes injustice, then whenever you have N_1, you must have injustice; whenever you have N_2, you must have injustice, and so on. Injustice therefore supervenes on each N_i. We can assume for the sake of argument that the necessity given by the realization relation is metaphysical necessity (Melnyk 2003). So the supervenience that realization gives us is supervenience with metaphysical necessity—in other words, the normative properties supervene on the descriptive properties with metaphysical necessity.

I argue that Brink's theory of normative properties does not give us an explanation of why Additive Reasons Supervenience expresses a metaphysical necessity.

Brink also holds that the multiple realizability of injustice is such that the various descriptive properties that constitute injustice in our world may not constitute injustice in other worlds. The example I gave was that the descriptive property of Nestle's monopolizing water distribution in poor areas of the world realizes injustice in our world, but it might not in worlds in which water wasn't a need. We can imagine a world in which water was like perfume: it's nice to have, but it's not a need, and so it's not a massive violation of human rights to monopolize water distribution systems and deny access to people that can't afford to pay. So although it is the case that, in our world, whenever you have Nestle monopolizing water distribution and access in poor areas of the world, you will have injustice, it's not the case that in every world, whenever you have Nestle monopolizing water distribution and access in poor areas of the world, you will have injustice. Injustice would then weakly but not strongly supervene on economic conditions of monopolization of water distribution and access.

Brink has a ready response to this objection. He can say that, although it's true that Nestle's monopolization of water doesn't realize injustice in every world, it's still true that Nestle's monopolization of water realizes injustice in every world that's exactly like ours in non-normative respects. So although injustice will not strongly supervene on economic conditions consisting of Nestle's monopolization of water access in poor countries, it will still strongly supervene on economic conditions consisting in Nestle's monopolization of water access in poor countries in worlds indiscernible from ours in all non-normative respects.

This argument succeeds at establishing a limited strong supervenience claim: that if something realizes injustice in our world, then its duplicate will realize injustice in worlds non-normatively indiscernible from ours. However, it does not succeed in establishing the broader kind of strong supervenience of the sort that we are looking for. As we saw in §2.6.3, the ecumenical formulation of normative supervenience was the view that any two individuals in *any two worlds* that were exactly alike in non-normative respects will also be such that these individuals will also be alike in all non-normative respects. As we put it earlier:

> *Full Strong Individual Additive Supervenience*: For any possible worlds w_1 and w_2 and any individuals i_1 in w_1 and i_2 in w_2, if i_1 and i_2 are intrinsically and extrinsically non-normatively indiscernible and w_1 and w_2 are non-normatively indiscernible, then, for any normative property P, i_1 instantiates P if and only if i_2 instantiates P.

What the earlier argument on behalf of Brink establishes is that any two worlds that are indiscernible from *our* world in all non-normative (purely descriptive) respects will be indiscernible with respect to the distribution of injustice. Ditto for the other normative properties and relations such as the favoring relation. So it establishes that any two worlds that are indiscernible from *our* world in all non-normative respects will be indiscernible in all normative respects. Since the kind of supervenience involved in realization holds with metaphysical necessity, the argument establishes that it's metaphysically necessary that any two worlds that are indiscernible from *our* world in all non-normative respects will be indiscernible with respect to the distribution of injustice.

However, this still leaves open the possibility that two worlds that are non-normatively indiscernible from each other but non-normatively *dis*cernible from ours can differ in

normative respects. For example, it leaves open the possibility that there are two non-normatively indiscernible worlds in which seaweed is sentient, but discernible in that in only one such world is the harvesting of seaweed to sell on the market unjust.

Brink can of course stipulate that these two non-normatively indiscernible sentient seaweed worlds have to be normatively indiscernible as well. We would then have strong supervenience. But such a stipulation would not give us what the irreduciblist looking for—namely, an explanation of why the two non-normatively indiscernible sentient seaweed worlds have to be normatively indiscernible.[18]

Causal regulation and contingent coextension

Brink's approach to establishing the strong supervenience of normative properties on non-normative properties with metaphysical necessity raises the concern that Brink's commitment to irreduciblism makes normative properties causally redundant and therefore unable to uniquely causally regulate our use of normative terms. This section will spell out this concern in more detail and its implications for explaining why the normative necessities are metaphysical necessities.

There is an argument against a posteriori varieties of psychological role functionalism by Jackson and Pettit (Jackson and Pettit 1993) that has been underappreciated in the philosophy of mind and, as far as I can tell, never been discussed in the literature on necessary a posteriori moral truths. Jackson's argument, extremely briefly, is that a posteriori functionalism in the philosophy of mind does not have the resources to distinguish between

[18]The argument I have hitherto given in this section is a modification and expansion of an argument given in (Sonderholm 2008).

psychological laws that hold of necessity and psychological laws that hold only contingently. The result is that the resulting functional theory will not generate the correct results when applied to counterfactual situations—or, if it did, it would do so only as a matter of improbable coincidence. In what follows, I give a similar argument, modified as appropriate for the moral case.[19]

Brink emphasizes that the reference of moral terms is causally regulated by "features of people, actions and institutions via [our] participation in an extended causal-historical chain linking past and present speakers of [our] language to the moral features of the world.... Moral appraisers can thus use moral language, such as "right" and "wrong", meaningfully, and to refer, provided that their use is part of the causal-historical chain" (Brink 2001, p. 163).[20] Brink also suspects, as I mentioned above, that the relevant features are features of causal systems whose effects conduce to the promotion and maintenance of human flourishing.

The identity of the referents of our moral terms is, on his view, an empirical question, and as far as I know, no philosopher has attempted to do the required empirical investigation herself. (Understandably so, for it would be quite the undertaking.) Suppose, however, that we grant Brink that the features that causally regulate our use of the moral terms are features of causal systems that conduce to the promotion and maintenance of human flourishing. We can further suppose with Brink that goodness is regulated by objects or states of affairs that are themselves constituents of human flourishing or that are causally disposed to produce

[19]In what follows, I also draw on the argumentative strategy of (Moffett 2010) and the literature on the exclusion problem for non-reductivism about mental states.

[20]See (Sayre-McCord 1997) for comments on the nature and relevance of the distinction between causal and non-causal regulation.

such states, and rightness is a property of actions that are so disposed. This is a rough exposition of a possible outcome of a moral functionalist empirical research program which I offer here in the absence of any more worked-out such view in the literature (as far as I know).

We now have two options: realizer functionalism or role functionalism about the referents of the moral terms. If Brink were a realizer moral functionalist, he could say that goodness, for example, just was whatever constellation of properties realized the goodness-causal role in the complex causal system that promotes and maintains human flourishing. Presumably this causal system would include, as parts, systems that promoted equitable distribution of water, Vitamin C and other vital nutrients to humans. This, however, generates worries analogous to the chauvinism worries that Block raised regarding realizer functionalism in the philosophy of mind. The flourishing and maintenance of Block's hypothetical silicon-based Martians, we can plausibly assume, will be promoted not by an equitable distribution of water and Vitamin C, but by equitable distributions of other nutrients that silicon-based Martians would require for their own maintenance and flourishing. If moral realizer functionalism identifies goodness with the distribution of water and Vitamin C, then it seems that causal processes that promote the maintenance and flourishing of members of the aforementioned alien species do not qualify as good. This, like in the mental chauvinism case, is intuitively the wrong result.

Brink, of course, is not a moral realizer functionalist but a moral role functionalist. It is open for a defender of such a position to hold that goodness is not to be identified with the property of bringing about an equitable distribution of water and Vitamin C, but with the property of bringing about an equitable distribution of whatever nutrients are needed to

promote and maintain the flourishing of sentient organisms. This is a higher-order property that will be shared both by the water-and-Vitamin-C-distribution processes that promote and maintain human flourishing, as well as the other distribution processes that would promote and maintain the flourishing of the rational Martians. Such a view avoids the "moral chauvinism" charge that I leveled at the above characterized moral realizer functionalist view.

As it turns out, there are an indefinite number of distinct, higher-order, functionally defined properties that are realized by the various lower-order causal processes that promote the maintenance and flourishing of human beings. Consider these principles:

1. Goodness is the property of being a process that (b) is causally disposed to produce *featherless bipedal* flourishing.
2. Goodness is the property of being a process that is causally disposed to produce *human* flourishing.
3. Goodness is the property of being a process that is causally disposed to produce *rational carbon-based lifeform* flourishing.
4. Goodness is the property of being a process that is causally disposed to produce *rational organism* flourishing.

Question: which of these role functional properties is the goodness property? All four are (contingently) coextensive properties. Any object or state of affairs with the property of being causally disposed to produce featherless bipedal flourishing also turns out, in our world, to possess the property of being causally disposed to produce human flourishing, rational organism flourishing, and vice versa. This is because the only featherless bipeds in

our world are also all and only the humans, all and only the rational carbon-based lifeforms, all and only the rational organisms, and so on. So we can't appeal to a difference in extension to help us adjudicate the question which of these distinct, higher-order properties is the property of goodness that's realized in our world by processes such as water distribution systems and so forth.

What most of us probably want to say is that goodness is the property of being a process that is causally disposed to produce sentient organism functioning. This gives us the intuitively correct result. But Brink is a causal regulation theorist about the reference conditions of "good" and other normative terms. He cannot rely on intuitions to figure out what "good" refers to. He has to do empirical investigation.

Brink himself alludes to this issue, although not explicitly as a potential problem for his view. In a footnote, he mentions that, instead of thinking of moral properties as those that play a causal role in the flourishing and maintenance of human organisms, they may be ones that play, instead, a causal role in the activities which are characteristic of sentient organisms (Brink 1984, p. 121). Brink gives no suggestion, as far as I can tell, as to how we can go about determining whether the causal systems are ones characteristic of sentient activity in general or human activity in particular. Later on, Brink does recognize this issue when he says that "it is often difficult to identify a strong supervenience base, because it is often difficult to be sure a set of base properties includes all the properties required literally to necessitate a set of supervening properties. So we often settle for identifying a weak supervenience base" (Brink 1989, p. 161).

Brink could maintain the following. Although it's true that we are unable to determine which of the contingently coextensive, higher-order properties is identical to goodness, we

can know, nevertheless, that, whichever such property goodness happens to be, the lower-order properties that realize goodness do so necessarily. Therefore the various lower-order properties that realize this higher-order property will be such that they constitute a strong supervenience base for goodness.

Brink cannot say exactly this since, as we saw before, he believes that it is possible that the properties that actually realize the normative properties in our world can fail to do so in other, non-nearby worlds. But Brink could say that this is not a problem, since these other, non-nearby worlds are not going to be non-normatively indiscernible. To put it another way, although this would function as a counterexample to something like Strong Individual Additive Supervenience that we saw in §2.6, it's not a counterexample to *Full* Strong Individual Additive Supervenience. Brink can therefore claim something like this:

> *Necessary Goodness Realization*: For any two possible worlds w_1 and w_2 and any causal systems x in w_1 and y in w_2, if w_1 and w_2 are non-normatively indiscernible and x and y are intrinsically and extrinsically non-normatively indiscernible, then, if x realizes goodness in w_1, then y realizes goodness in w_2.

Since realization relations are metaphysically necessary, Brink would then have an explanation of why goodness supervenes on non-normative reality with metaphysical necessity, even though we can't in principle know which of the higher-order properties it is that goodness happens to be.[21]

If Brink were not a causal regulation theorist about goodness, I think this response would work. However, I think that Necessary Goodness Regulation is problematic when conjoined

[21] This argumentative strategy will not work for Brink if the argument of the previous section is sound. In this section, I set those concerns aside.

with Brink's causal regulation view of "good". As stated above, Brink accepts the following claim:

> Each moral term *t* rigidly designates the [non-normative] property *N* that uniquely causally regulates the use of *t* by humans. (Brink 2001, p. 165)

Applying this to the case of goodness, we get the result that "good" rigidly designates the non-normative property that uniquely causally regulates the use of "good" by humans. However, we also saw above that there are an indefinite number of higher-order properties that contingently coextend with the good things: the property of being a process that is causally disposed to produce human flourishing; or the property of being a process that is causally disposed to produce sentient organism flourishing, and so on. There is a worry here that contingently coextensive properties will be either *causally redundant* or *causally inefficacious* and thus cannot uniquely causally regulate the use of "good" by humans.

To see this worry, we can look at another example from philosophy of mind. Jaegwon Kim argued that non-reductive physicalists in the philosophy of mind face what he called the *exclusion problem*: that, if non-reductivism about mental events is true, then mental states and events are either causally efficacious or causally redundant (Kim 1989), (Kim 1998), (Kim 2005). The exclusion problem is generated from the tension present within the following three claims:

1. *Physical Causal Closure*: All physical effects have sufficient physical causes.
2. *Mental Causation*: Some physical events have mental causes.
3. *Non-Reductivism about Mental Causes*: Mental causes are numerically distinct from physical causes.

Since all physical effects have sufficient physical causes, any mental cause of a physical effect will be causally redundant in the sense that it overdetermines any physical effect that it causes. If one furthermore accepts that there are no causally redundant causes, then mental states and events cannot be causes of physical effects.

Here is a modification of Physical Causal Closure and Non-Reductivism about Mental Causes that are tailored to the case of causally regulated normative terms:

1. *First-Order Causal Closure*: For any causally regulated term t, there is a property or set of properties that constitute a sufficient first-order cause of any successfully denoting use of t.
2. *Causal Regulation of "Good"*: All uses of "good" are uniquely causally regulated by goodness.
3. *Irreduciblism about Goodness*: Goodness is not a first-order property.

(2) and (3) are aspects of Brink's view. Let's set the truth or falsity of (1) aside just briefly. If, by (1), there is a first-order cause of any successfully denoting use of "good", then "good" is causally regulated by this first-order cause. But, by (2), all uses of "good" are *uniquely* causally regulated by goodness. Goodness would therefore have to be whatever first-order cause it was that causally regulates any successfully denoting use of "good". However, by (3), goodness is not a first-order property and therefore not numerically identical to the first-order cause of successfully denoting uses of "good". So goodness must be causally redundant, which it can't be, since, by (2), goodness is supposed to be the unique property that causally regulates all successfully denoting uses of "good".

The irreduciblist has three options here. He can give up (3): the claim that goodness is identical to a higher-order role property. He can instead claim that it is identical to a first-order realizer property. In this case he becomes a reduciblist and therefore can't give the irreduciblists a route to explaining why normative supervenience is metaphysically necessary. He also runs into the moral chauvinism worries discussed above.

The irreduciblist could give up (2), which is the causal regulation theory of the reference of normative terms. If he does this, however, then he can't use the causal regulation theory to explain why normative supervenience is metaphysically necessary.

Therefore, irreduciblists who want to use the causal regulation theory of reference to explain why normative supervenience is metaphysically necessary have to hold that (1) is false. Brink would likely hold that (1) is false. Consider again from §3.3.1 the two ways that causally regulated terms get their reference conditions. Causally regulated predicate terms can get their reference conditions through ostension, in which case the linguistic community gestures at some samples and stipulates through a kind of speech act that the term is to apply to the property that these samples have in common. Applying this to the case of goodness, we can imagine that the initial linguistic community took every sample of goodness-instantiating things available to them and gestured at the entire set. Goodness, however, is multiply realizable, so there is not going to be a single first-order realizer property that all the samples of goodness share in common.[22] Causally regulated predicate terms can also acquire reference by being associated with a particular description and gaining their reference by being stipulated to be the property common to individuals that satisfy

[22]There would be such a property if Brink allowed disjunctive properties, but Brink does not allow for disjunctive properties.

that description. However, again, goodness is multiply realizable. So there is not going to be a single first-order realizer property that all the samples of goodness share in common. Brink can therefore deny (1).

However, the worry about causal redundancy comes up again even if we limit ourselves to considering the causal impact of the higher-order role properties. I mentioned earlier that there are an indefinite number of higher-order role properties that all good individuals contingently share in common. Two examples I gave of these contingently coextensive, higher-order properties were the properties of being such that these individuals promoted the flourishing of human beings, and of being such that these individuals promoted the flourishing of rational organisms. Suppose the causal regulation theory of "good" is true. The initial linguistic community therefore assembled the sample of good things, either through ostension or through some sort of initial description, and stipulated that goodness was the property that all these samples had in common. Since goodness is multiply realizable, there is no first-order property that the initial baptism could pick out. The samples do, however, have in common an indefinite number of numerically distinct, higher-order role properties. So here is a question: which of these does the initial baptism pick out?

If the initial baptism picks out none of them, then there is no property that causally regulates "good". "Good" would then have no reference, and the causal reference approach would just give us error theory rather than an irreduciblist variety of normative realism. If the initial baptism picks out all of them, we have a contradiction, since it's metaphysically and conceptually impossible for a property to be numerically identical to multiple numerically distinct properties.

So the initial baptism has to pick out a single one of the indefinitely many properties

that all the samples share in common. However, this is also impossible. The causal regulation theory of reference says that any successfully denoting use of a causally regulated term t successfully denotes its referent in virtue of a complex causal chain of events that traces back through the initial baptism to the referent itself. It is the referent of t itself that causes a use of t, through the initial baptism and the subsequent complex chain of causes, to successfully denote its referent. However, in the case of role properties, there is an intermediary step between the initial baptism and the role property that the baptism picks out. That intermediary step is whatever realizer properties of the sample objects realize the role property. Role properties are higher-order properties of the lower-order realizer properties instantiated by the sample objects, so in order to get to the higher-order properties, you need to go through the lower-order properties first. However, since the lower-order properties in question are lower-order properties of indefinitely many higher-order properties, the lower-order properties create a kind of "causal bottleneck" that prevents the initial baptism from "landing" on one unique role property. If it does manage to land on one unique role property by means of some causal process, it is not at all clear what such a causal process would look like.[23]

Importantly, the problem of identifying which higher-order properties causally regulate our terms is not a problem for the standard cases of natural kind terms and proper names.

[23]One possibility is that the initial baptism did not simply baptize the term "good", but rather baptized a number of distinct, homographic terms "good$_1$", "good$_2$, ... "good$_n$" ..., one for each of the indefinitely many contingently coextensive higher-order role properties that the good things share. We could then say that it is vague which of these terms we are using when we say "good" on any given occasion. If the type of vagueness were epistemic vagueness a la (Williamson 1996), then each use of "good" would pick out a unique role property and irreduciblists would be closer to having an explanation for why normative supervenience is metaphysically necessary. This approach bears certain similarities to the theory of reference presented in (LaPorte 2004). This may be a promising avenue for future research into solutions for the causal regulation theory of the reference of normative terms. However, I will not explore this idea further here, hence its relegation to this lowly footnote.

The reason for this is that these standard cases are not cases of causal regulation by higher-order role properties. The natural kind cases are cases of causal regulation by the first-order properties that are essential to the natural kind that was implicated in the initial baptism. The proper name cases are cases of causal regulation by individuals. So my comments here are not meant to be a criticism of causal regulation theories of reference in general, but only for causal regulation theories of predicate terms that take the properties that causally regulate these terms to be some higher-order property that contingently coextends with numerically distinct higher-order properties.

4.3 Neo-essentialism

There is a third approach to establishing metaphysical necessity that we should consider alongside the conceivability and causal regulation approaches that we have discussed above: Kit Fine's essentialist account of modality (Fine 1994). In the next two paragraphs I will give an extremely brief overview of Fine's account in order to set the stage for an argument that, on his account of metaphysical modality, we actually have reason to reject irreduciblism.

Neo-essentialist metaphysical necessity

An individual's essential properties are the properties that make individuals what they are. They are the properties that are "non-accidental to" that individual; or, to put it slightly differently, if *P* is an essential property, then it is in the nature of that individual that it has *P*. On the influential modal characterization of essential properties, an individual's essential

properties are those and only those properties that an individual must possess in order to be the individual that it is. Fine famously objects to modal characterizations of essential properties. If Socrates exists, then necessarily the singleton containing Socrates as its only member also exists. Intuitively, we do not want to say that it is essential to Socrates that he belong to the singleton. Similarly, if Socrates exists, it is necessary that Socrates exist. But we do not want to say that Socrates' existence is essential to Socrates. Socrates is not a necessary being.

Rejecting the modal analysis of essence raises a question as to how to characterize the relation between modality and essence. Here is one way. If a certain property—say, rationality (or rational potential, as the Scholastics like to say)—is essential to Socrates, then, necessarily, Socrates will be rational (or potentially rational) if he exists at all. Fine argues that this gives rise to a number of sets of necessary truths, each of which are true in virtue of an underlying essence of something. "Socrates is rational" is necessarily true in virtue of the essence of Socrates, "Socrates belongs to the singleton Socrates" is necessarily true in virtue of the essence of the singleton, and so forth. Fine suggests that metaphysical necessity can be understood as the union of all these truths that are necessary in virtue of object types:

> [W]e should view metaphysical necessity as a special case of essence. For each class of objects, be they concepts or individuals or entities of some other kind, will give rise to its own domain of necessary truths, the truths which flow from the nature of the objects in question. The metaphysically necessary truths can then be identified with the propositions which are true in virtue of the nature of all objects whatever. (ibid., p. 9)

By claiming that the metaphysically necessary truths can be identified with the propositions

which are true in virtue of the nature of all objects whatsoever, Fine doesn't mean that the only metaphysically necessary truths are truths that are essential to every object whatsoever. He means that the set of metaphysically necessary truths are the set of truths such that, for each truth T in the set, there is an object x such that T is true of x in virtue of x's essence. For example, it is metaphysically necessary that Socrates, if Socrates exists, is rational, even if "if Socrates exists, Socrates is rational" is not an essential truth of, say, the number 2. Furthermore, Fine uses the term "object" in a quite capacious way, to include not only individuals with properties but also concepts.

Neo-essentialism and irreduciblism

Supervenience and necessary coextension

As it turns out, our Full Strong Individual Additive Supervenience claim entails that for every normative property—e.g., moral rightness—there exists a non-normative property *D** that is necessarily coextensive with it:

Necessary R-coextension: $\Box\forall x[R(x) \leftrightarrow D^*(x)]$.

A similar claim will be true of any normative property. So, if goodness is a normative property, then, given Full Strong Individual Additive Supervenience, there will also be a fully non-normative property that is necessarily coextensive with goodness, and so on.

The way to construct the property involves using a Ramsification procedure that was elaborated by Kim in his quest to construct psychophysical bridge laws and was applied to the case of normative properties by Jackson.

First, we take the set of all individuals in all possible worlds $x_1, x_2, ..., x_n...$ such that $R(x_i)$, where *R* denotes moral rightness. Let $w_1, w_2, ..., w_n...$ be the worlds of $x_1, x_2, ..., x_n...$ respectively. Second, for each x_i, take every intrinsic and extrinsic non-normative property that x_i has and construct a property D_i which is just a conjunction of all of these intrinsic properties, extrinsic properties, and the property of being in a world that's non-normatively indiscernible to w_i. Call any such property a *complete description property*. Jackson assumes that every right individual has non-normative properties, but this assumption is wholly uncontroversial as far as I can tell.

Third, we take the disjunction of complete description properties $D_1, D_2, ..., D_n...$, which gives us $D_1 \vee D_2 \vee ...D_n....$ Call this *D**.[24]

It follows that $\Box\forall x[R(x) \leftrightarrow D^*(x)]$.

I'll start with the left-to-right direction: $\Box\forall x[R(x) \rightarrow D^*(x)]$. $\forall x[R(x) \rightarrow D(x)]$ because, necessarily, every right individual has non-normative properties and is located in a world with non-normative features. The procedure then includes any D_i such that $R(x_i)$ as a disjunct of D^*. Therefore, because of the truth conditions of disjunction, $\forall x[R(x) \rightarrow D^*(x)]$. $\Box\forall x[R(x) \rightarrow D^*(x)]$ because we ran the construction procedure for all morally right individuals in all possible worlds.

Now for the right-to-left direction $\Box\forall x[D^*(x) \rightarrow R(x)]$. $D^*(x)$ is a big disjunction of complete descriptive properties $D_1, D_2, ..., D_n...$ Therefore, by the truth conditions of disjunction, any x such that $D^*(x)$ will also be such that there is a property D_i in D^* such that $D_i(x)$. Now, Jackson's procedure included the complete description property D_i as a

[24] I mentioned earlier in §3.3.2.2 that disjunctive properties are controversial. I discuss some of these issues in §3.3.3.3 below.

disjunct of D^* *only if* $\exists x[D_i(x) \wedge R(x)]$. Therefore, any x such that $D^*(x)$ will also be such that there is a property D_i in D^* such that $D_i(x) \wedge \exists y[D_i(y) \wedge R(y)]$.

D_i is a complete description property. It is the property of being intrinsically and extrinsically non-normatively indiscernible to x_i in a world non-normatively indiscernible to w_i. Let's compare that to Full Strong Individual Additive Supervenience:

> *Full Strong Individual Additive Supervenience*: For any possible worlds w_1 and w_2 and any individuals i_1 in w_1 and i_2 in w_2, if i_1 and i_2 are intrinsically and extrinsically non-normatively indiscernible and w_1 and w_2 are non-normatively indiscernible, then, for any normative property P, i_1 instantiates P if and only if i_2 instantiates P.

Full Strong Individual Additive Supervenience just says that if two individuals are intrinsically and extrinsically non-normatively indiscernible and are in non-normatively indiscernible worlds, then these two individuals will both have or lack all the same normative properties. But if two individuals are intrinsically and extrinsically non-normatively indiscernible and are in non-normatively indiscernible worlds, this is just to say that these two individuals have the same complete description properties. That's what complete description properties are. So Full Strong Individual Additive Supervenience tells us that if two individuals have the same complete description properties, then they will also have and lack all the same normative properties. Therefore, if it's true that any x such that $D^*(x)$ will also be such that there is a property D_i in D^* such that $D_i(x) \wedge \exists y[D_i(y) \wedge R(y)]$, then Full Strong Individual Additive Supervenience tells us that any x such that $D^*(x)$ will also be such that $R(x)$. $\forall x[D^*(x) \rightarrow R(x)]$. $\Box\forall x[D^*(x) \rightarrow R(x)]$ because we ran the construction procedure for all morally right individuals in all possible worlds.

Jackson's procedure also works with sets of individuals and normative relations. So, given Full Strong Individual Additive Supervenience, we could also use the procedure to construct a fully non-normative property that was necessarily coextensive with the favoring relation.

Irreduciblism, neo-essentialism, and necessary coextension

With that done, we are in a position to see why neo-essentialist accounts of metaphysical modality don't provide irreduciblists with an explanation of why Additive Reasons Supervenience and other additive varieties of supervenience do not express metaphysical necessities.

The argument I give here is essentially one that Gideon Rosen makes in unpublished work. However, the conclusions Rosen and I draw are distinct. Rosen makes this argument in order to give us reasons to call normative supervenience into question. My conclusion is that the irreduciblist cannot appeal to neo-essentialist accounts of modality to explain why Additive Reasons Supervenience does not express a metaphysical necessity. Furthermore, the extension of Rosen's argument to respond to someone who denies the existence of disjunctive properties is my own work.

Let's assume that Jackson's construction procedure generated the following claims:

> *Pleasure Promotion entails Rightness*: Necessarily, for all individuals x, if x promotes pleasure over pain, then x is morally right; and
>
> *Rightness entails Pleasure Promotion*: Necessarily, for all individuals x, if x is morally right, then x promotes pleasure over pain;

where, in both cases, the promotion of pleasure over pain is thoroughly non-normative. The reduciblist who accepts both claims claims that moral rightness is identical to the promotion of pleasure over pain. Since the reduciblist thinks that moral rightness is identical to the promotion of pleasure over pain, the reduciblist will accept that being morally right consists in nothing over and above the promotion of pleasure over pain.[25] Jackson's construction procedure forces the irreduciblist to agree with the reduciblist that moral rightness is necessarily coextensive with the promotion of pleasure over pain, but the irreduciblist will claim that moral rightness consists in something over and above the promotion of pleasure over pain.

Suppose that the neo-essentialist account of modality were true. Then irreduciblism about moral rightness must be false. Here is an argument from contradiction to show that this is the case.

On the neo-essentialist account of modality, it is a metaphysical necessity that Socrates belong to the singleton (if the singleton exists), because it is part of the essence of the singleton Socrates that Socrates belong to the singleton Socrates. In other words, what it is to be the singleton Socrates is for Socrates to belong to it.

Similarly, it would be a metaphysical necessity that actions that promote pleasure over pain were morally right just if it were part of the essence of pleasure promotion that actions that promote pleasure promotion be morally right. In other words, if it were a metaphysical necessity that actions that promote pleasure over pain were morally right, then what it would

[25]The reduciblist, of course, will not think that the promotion of pleasure over pain is non-normative, since, after all, the reduciblist thinks that moral rightness just is the promotion of pleasure over pain, and that moral rightness is normative. Reduciblists about normativity are not eliminativists about normativity. The reduciblist would say that the promotion of pleasure over pain is not a non-normative property but rather a normative property that is also a descriptive property.

be for an action to be an action of promoting pleasure over pain is for that action to be morally right.

But this is to say that it is part of the essence of promoting pleasure over pain that the promotion of pleasure over pain be normative. If this were true, then we have a contradiction. For the promotion of pleasure over pain was ex hypothesi thoroughly non-normative. Jackson's construction procedure was a procedure that constructed a thoroughly non-normative property that necessarily coextended with moral rightness. Full Strong Individual Additive Supervenience guarantees that there is such a property. Therefore Jackson's construction procedure could not generate the promotion of pleasure over pain as a property that's necessarily coextensive with moral rightness.

The upshot of this argument is that, if the irreduciblist accepts the neo-essentialist account of modality, she cannot conclude that it's metaphysically necessary that actions that promote pleasure over pain are morally right. If she cannot conclude that it's metaphysically necessary that actions that promote pleasure over pain are morally right, then a forteriori she cannot conclude that it's metaphysically necessary that actions promote pleasure over pain if and only if they are morally right. The normative necessity that actions promote pleasure over pain if and only if they are morally right therefore fails to be a metaphysical necessity.

The irreduciblist could claim that it is essential to morally right actions that morally right actions promote pleasure over pain. The irreduciblist is committed to the claim that moral rightness consists in something more than the promotion of pleasure over pain, but she can still claim that moral rightness consists at least partly in the promotion of pleasure over pain. This claim would be consistent with the irreduciblist view. However, if it were

essential to morally right actions that morally right actions promote pleasure over pain, then this truth would be true in virtue of the nature of moral rightness. It would establish, in other words, that it were essential to the normative that morally right actions promoted pleasure over pain. This gives us a necessary truth, but the kind of necessary truth it gives us is a normative necessity. It would not do anything to show that this truth, which is a normative necessity, is also a metaphysical necessity.

An aside on disjunctive properties

Jackson's construction procedure assumes that the set of properties is closed under disjunction. In other words, it assumes that, for all P and Q, if P and Q are properties, then $P \vee Q$ is a property. I mentioned in the previous section §3.3.2.2 on causal regulation approaches to metaphysical necessity that this disjunctive closure principle is controversial. However, in the example I am using here of rightness as pleasure promotion, no disjunctive properties are implicated. So the argument I have made succeeds irrespective of whether one accepts or denies the existence of disjunctive properties.

It's true that Jackson's construction procedure does not guarantee the existence of a necessarily coextensive property if we do not accept that properties are closed under disjunction. However, if there is not a fully non-normative property that is necessarily coextensive with rightness, then it is even less plausible that it is in the essence of any non-normative property that objects instantiating that property be normative. After all, if rightness is not necessarily coextensive with pleasure promotion, then there will either be an instance of pleasure promotion that is not an instance of rightness, or an instance of rightness that is

not an instance of pleasure promotion. If there is a possible instance of pleasure promotion that is not an instance of rightness, then it can't be essential to pleasure promotion—it cannot be part of what it is to be pleasure promotion—that pleasure promotion be right. If it were, then any action would have to be right in order to be an instance of pleasure promotion. That's part of what it means for a property to be an essential property. There would therefore could not be any instance of pleasure promotion that were not right. It would be metaphysically impossible.

Now suppose that there were an instance of rightness that were not an instance of pleasure promotion. You tell the Nazis at your door that there are Jews in your basement, and this, we'll assume is morally right even though your honesty gets you and all the Jews killed. Either every truth-telling action is necessarily right, or it is not. If it is not, then it will not be essential to actions of truth-telling that they be right, for the reasons mentioned above. If it is, then we can generate the following two principles:

> *Pleasure Promotion entails Rightness*: Necessarily, for all individuals x, if x promotes pleasure over pain, then x is morally right and
>
> *Honesty entails Rightness*: Necessarily, for all individuals x, if x is an instance of truth-telling, then x is morally right.

Now, "being such as to promote pleasure over pain" and "telling the truth" are non-normative properties, or they are not. If they are not, then it's not essential to either property that actions with these properties be right, which, on the neo-essentialist picture, means that the necessity is not metaphysical necessity. If they are, then they will also count as normative properties and will not support an additive supervenience claim, for the reasons

discussed in the earlier section.

Chapter 5

Saving Irreducibility

In the previous chapters 3 and 4, I argued that the prospects for an irreduciblist explanation of why normative reality strongly supervenes on non-normative reality (or, in other words, that it supervenes additively) are dim. In this chapter, I argue that the irreduciblist should give up on additive supervenience in favor of what I call transformative supervenience. They can give up on additive supervenience by either denying it outright or by accepting Bilgrami's claim that additive supervenience is incoherent. By accepting transformative supervenience, irreduciblists gain the resources to explain why the transformative supervenience relation obtains while simultaneously doing justice to the motivations for accepting a supervenience claim in the first place and generating an attractive and novel theory[1] of the relation between normativity and the world. The result, however, will entail a significant departure from most other commitments of the reasons fundamentalist view described in §2.

The view that I defend in this chapter is a variety of what Susan Hurley calls *non-centralism* about thin properties (Hurley 1989, pp. 9-29). If you're a *centralist* about thin concepts and properties, you think that thin concepts are conceptually independent of thick concepts, and that thin properties are explanatarily prior to thick properties. The conceptual

[1] I should clarify that it's presumptuous to describe what I give in this chapter as a "theory". It's more like a sketch of a theory, or a direction that theorizing could go.

independence here should be understood as follows: A is "conceptually prior to B just if one can correctly grasp A without grasping B. The centralist will also hold that the thin properties are constitutively more fundamental to the thick properties in that part of what it is for a thick property to be present is for a thin property to be present as well.[2]

If you're a non-centralist about thin properties, you deny centralism about thin properties. There are two ways of being a non-centralist. You can hold, as Hurley does, that the thick and thin concepts and properties are conceptually and constitutively *inter*-dependent. Or you can hold, as I will, that the thick concepts and properties are conceptually independent and of and constitutively more fundamental to thin concepts and thin properties. I am, however, not opposed to Hurley's interdependence view and will make some remarks about this view later on in this chapter.

Here is the list of claims that will feature in the theory of normativity that I will develop across this chapter. Note the similarities and differences with the components of reasons fundamentalism.

> *Reasons Thin Centrality*: For all facts F, F is a *thin* normative fact only if F is a reasons fact or F embeds a reasons fact.
>
> *Extended Reasons Relationality*: Any reason fact is normative in virtue of its standing in the favoring relation with a *quintuple* consisting of a fact, an agent, an action, a circumstance, a contrast class of actions, and a feature of the action with respect to which the fact is a reason.
>
> *Answers Account of Reasons*: Reasons are *partial correct answers* to questions why

[2]This description is is accurate to the centralist position that Hurley describes, but she doesn't quite put things like this.

an agent should perform an action rather than not, in a particular circumstance, with respect to a thick property that the action would promote.

Reasons Irreducibility: The favoring relation is not normatively masked.

Transformative Reasons Supervenience (*TRS*): For any possible worlds w_1 and w_2 and any reasons-apt tuples T_1 in w_1 and T_2 in w_2, if T_1 and T_2 are intrinsically and extrinsically *descriptively* indiscernible and w_1 and w_2 are *descriptively* indiscernible, then T_1 instantiates the favoring relation if and only if T_2 instantiates the favoring relation.

I ignore the *Non-Humean Normative Motivation* conjunct because, as I mentioned in §2.7.2, the truth or falsity of this conjunct is largely independent of the two primary topics of this dissertation: the role that the favoring relation plays in the structure of normative reality, and the nature of the relation between normative reality and non-normative reality.

There are a lot of changes between these claims and the components of reasons fundamentalism that I stated at the beginning of §2. Of the five claims here, four of them are modified versions of conjuncts that I claimed constituted reasons fundamentalism at the beginning of §2. The only one of the five characterizing features of reasons fundamentalism that my proposed theory preserves in its original form is Irreducibility.

Reasons Thin Centrality limits the *Reasons Centrality* claim from §2 only to thin normative facts. Thin normative facts are instantiations of thin normative properties. I characterized thin normative properties as normative properties that are not descriptive properties. Normative properties are properties whose instantiations have direct consequences for the way the world ought to be or could ought to be. Descriptive properties are properties whose

instantiation makes or would make the world the way it is or could be. "Thick" properties are properties that are both descriptive and normative: their instantiations both make or could make the world the way it is or could be, and their instantiations also have direct consequences for the way the world ought to be or could ought to be.

I am attracted to Reasons Thin Centrality because it is an implication of Scanlon's buck-passing view of thin properties. I mentioned earlier that Scanlon has influentially argued that thin properties such as goodness, value and wrongness are purely formal, higher-order properties of objects and actions that have lower-order properties that provide reasons of the goodness kind, the value kind, or the wrongness kind. I find Scanlon's arguments for this view in (Scanlon 1998) to be convincing.[3] However, this dissertation will not offer any arguments for Reasons Thin Centrality.[4] It will only describe what the theory I sketch here would look like if Reasons Thin Centrality were also incorporated. I give this description in §5.6, at which point the other aspects of my theoretical sketch will be in place.

Extended Reasons Relationality states that reasons-apt tuples (in other words, those tuples that are apt for instantiating the reasons relation) are not four-place tuples but rather six-place tuples. The success of my argument for this also turns out to generate an interesting response to the reasons fundamentalist's commitment to *Reasons Primitivity*. According to the *Answers Account of Reasons* component of the theory of normativity that I am defending, our concept of a reason does indeed admit of a non-trivial analysis in terms

[3]Scanlon of course adopts a buck-passing view of all normative properties other than the favoring relation. I do not follow him in this, since, as I argued, I believe that this commits him to Additive Reasons Supervenience, which I argued generates problems for explaining why the normative supervenience relation that he accepts holds with metaphysical necessity. I only follow him in adopting the buck-passing view of thin normative properties.

[4]If the reader wants such arguments, the place to look is chapters 1 and 2 of (Scanlon 1998).

of answers to questions why an agent's performing an action in a particular circumstance will promote a thick property. *Answers Account of Reasons* is meant to be both an analysis of our concept of a reason as well as a real definition (i.e. an account of what reasons are essentially).

Finally, Transformative Reasons Supervenience (TRS) formulates the supervenience of reasons as the supervenience of the favoring relation on the distribution of descriptive properties. TRS is a strictly weaker supervenience claim than Additive Reasons Supervenience (ARS). TRS is strictly weaker because the set of descriptive properties that feature in my statement of TRS includes the set of thick properties along with the set of non-normative (purely descriptive) properties. The class of descriptive properties is strictly broader than the class of non-normative (purely descriptive) properties. Therefore, anyone who accepts ARS is also accepting a stronger version of TRS. One can, however, accept TRS without accepting ARS since one can hold, in principle, that differences among the distribution of thick properties in a world are also relevant to determining whether two reasons-apt tuples are relevantly similar for the purpose of evaluating the supervenience claim.

I worry that this is not clear, so here is a toy example to illustrate the point. Suppose Lee says that SAT scores are a test of intelligence—in other words, that the score that you get when you take the SAT is fixed by your intelligence. On Lee's view, since your SAT score is totally fixed by your intelligence, there is no room for something like extensive test prep to make any impact on what score you receive. Mary disagrees. She thinks that, yes, intelligence does play a role, but that test prep also makes a significant difference as to how well you perform on the SAT. Mary, for instance, thinks that two people with equal intelligences could still achieve unequal SAT scores if one person had access to a lot of

test prep services that the other person did not. Mary's view is more permissive than Lee's in the sense that she thinks it's possible, and Lee does not, that two people have the same intelligence but differing SAT scores. When I say that TRS is more permissive than ARS, I mean something similar: that it's open for the advocate of TRS to hold that two worlds that are non-normatively indiscernible can be normatively discernible due to a difference in the distribution of thick properties. This possibility is not open for the advocate of ARS.

After talking more about what thick properties are in §5.1, I will start this chapter with a discussion in §5.2 and §5.3 of why TRS can help an irreduciblist avoid the concerns about Brute Entailment and Brute Normativity that I raised in §3.1 for reasons fundamentalist views. I'll then move on in §5.4 and §5.5 to develop arguments for Extended Reasons Relationality, and Answers Account of Reasons. I will not be offering arguments for Reasons Thin Centrality, but I will give some indications in §5.6 as to what a theory that accepted Extended Reasons Relationality, Answers Account of Reasons, and Transformative Reasons Supervenience would look like if it also accepted Reasons Thin Centrality.

5.1 What are thick properties?

The term "thick concept" traces to (Williams 1985, p. 143). Previously, Foot and Hare had discussed thick concepts, but not under that label. Williams' discussion of thick concepts appears in the context of his discussion of the fact-value distinction and the central role that this distinction played in the moral philosophy of that time.

> What has happened is that the theorists have brought the fact-value distinction to language rather than finding it revealed there. What they have found are a lot of those "thicker" or more specific ethical notions I have already referred to,

> such as treachery and promise and brutality and courage, which seem to express a union of fact and value. The way these notions are applied is determined by what the world is like (for instance, by how someone has behaved), and yet, at the same time, their application usually involves a certain valuation of the situation, of persons or actions. Moreover, they usually (though not necessarily directly) provide reasons for action. (ibid., pp. 129-130)

Terms like "brutality" and "courage", on this picture, express thick concepts, which are concepts composed of both "fact" content and normative content. Similarly, we can say that a thick *property* is a property that is both a "fact" property and a normative property. The "fact property" locution, though, is misleading. Normative properties would trivially be "fact properties" in the sense that there are normative facts, and that normative properties partially constitute them. Sarah is good. That's a fact. Goodness is trivially a "fact property" just in virtue that it is a property, and as such can partly constitute facts, such as the fact that Sarah is good and so on. Instead of characterizing thick properties as properties that are both fact properties and normative properties, it is less misleading to characterize thick properties as properties that are both *descriptive* and normative. Descriptive properties are just properties that partially constitute facts about how the world is or could be, whereas normative properties are properties that partially constitute facts about how the world ought to be or could ought to be.

It is important to distinguish between thick concepts and thick properties in part because we are error theorists about many thick concepts. Take Gibbard's example term "gopa", which is a term in the language of a fictional indigenous people that Gibbard introduces in a thought experiment aimed at discovering the semantics of thick concepts (Blackburn and Gibbard 1992, pp. 267-268). "Gopa" is an adjective used by the fictional people for glorious acts of killing outsiders. It's a thought experiment, so we will just take Gibbard's word that

this is the appropriate translation to the term in the fictional language. We don't think that killing outsiders is glorious, so we don't think acts are gopa. We are, in other words, error theorists about gopahood. Additionally, the distinction matters because, as Blackburn notes (ibid.), the most compelling examples of thick concepts are ethnic pejoratives. Ethnic pejoratives are terms that express a concept of being of a certain ethnicity and somehow morally inferior for being so. We are error theorists about ethnically pejorative statements because we do not think that there are any people who are morally inferior in virtue of their ethnicity. Williams is therefore correct to claim that thick ethical notions only *usually* provide reasons for action, since ethical pejoratives don't denote properties and therefore, a forteriori, cannot denote properties that constitute facts that provide reasons for action.

Thick normative concepts are commonly distinguished from normative concepts that lack descriptive content. These purely normative concepts are called "thin" normative concepts. Purely normative concepts are called "thin" concepts because they lack the "thickness" of descriptive content; there is no descriptive content that is *just theirs*.[5] They are generally thought to include the terms "good" and "right", which can be consistently predicated to objects or actions independently of whatever other descriptive properties those objects or actions possess. If we say that Sarah instantiates a thick property such as courage, we make certain descriptive claims of Sarah–for instance, that she is disposed to try to control her fear responses in the face of perceived danger. We would contradict ourselves to say that Sarah was courageous and never tried to avoid running away from danger. However,

[5]There is an interesting relation between this claim and Moore's claim, discussed earlier in the "Jettisoning Naturalism" section in the chapter 2 discussion of irreducibility and naturalism, that a "natural" property is a property that is visible. Moore may be operating under an assumption that meaningful content, in other words content that we can say something about, is only given by visible or otherwise empirically detectible properties. Goodness being an invisible property leads Moore to conclude that it is therefore a "non-natural" property, or a property that we can't say anything meaningful about.

there is no contradiction generated by saying that Sarah is good and never tries to avoid running away from danger.[6] This is because the concept of goodness, unlike the concept of courage, does not contain descriptive content that can generate inconsistencies such as the one that I have described in the case of courage. Similarly, if we say that an action is brutal, we can't also consistently say that the action pleases many people and harms no one. But we could say without conceptual contradiction that an action is wrong and that it pleases many people while harming no one. We could still be wrong about this, but we would not be guilty of a *conceptual contradiction* if we were to say it. This is because the concept of brutality partly includes descriptive content such as harming some sentient being, but wrongness does not include any such descriptive content.

It's true that there would be a conceptual mistake involved in saying that Sarah was right. Rightness can only be a property of actions, and Sarah is not an action. There is a sense, then, that even the thinnest ethical concepts entail some constraints about the kinds of entities that they can be consistently predicated of. Unfortunately, there is hardly any discussion of this issue in the literature. So here is my own attempt to understand what the thick/thin normative concept distinction amounts to.

Thin concepts are not concepts that can be exceptionlessly consistently predicated of anything whatsoever. They are concepts that have no descriptive content that could conflict with the descriptive features of an individual beyond what you can read off of the underlying metaphysics given by the grammar of the terms denoting that individual. "Sarah is morally

[6]I understand courage as a commitment to control one's emotional responses in the face of perceived danger rather than a disposition to control one's fear responses. Given this understanding of courage, it is possible for a person to have a commitment to courage and yet lack the disposition to courageous behavior. I understand a commitment to X in a situation Y to be a first-order mental state that an agent has if and only if that agent is disposed to *try to* X in Y. I owe these clarifications and general view to Carol Rovane and Akeel Bilgrami.

right" generates a contradiction because we know from the grammar of this sentence that Sarah is a person, and persons cannot be morally right. "Sarah is morally good" does not generate a contradiction merely from the grammar, since persons are among the types of entities that can instantiate moral goodness. The possibility of these sorts of contradictions is irrelevant to the determination of whether a concept is thick or thin.

The difference between thick and thin concepts has to do, rather, with whether a conceptual contradiction could immediately result from spelling out the descriptive nature of the individual instantiating the normative property in more detail. If we said that Sarah is courageous, this could immediately generate a conceptual contradiction if Sarah turned out to always dissociate from fearful situations. We could say "But Sarah is not courageous, since she dissociates in the presence of fearful situations". However, no purely descriptive fact that could be true of Sarah would immediately generate a conceptual contradiction from the assertion that Sarah is morally good. It might, of course, implicate that given further assumptions about moral goodness. We could discover that Sarah profits off of predatory lending practices, in which case Sarah would not be morally good. But there is nothing about the concept of moral goodness in itself that entails this. You would need additional background theory about the nature of moral goodness. Much more would need to be said about the approach to demarcating thick and thin normative concepts that I give here, but hopefully these remarks do a little more to clarify what I take the salient distinction between the thick and thin normative concepts happens to be.

5.2 Transformative Supervenience and Brute Entailment

Here is Transformative Reasons Supervenience again:

> *Transformative Reasons Supervenience* (TRS): For any possible worlds w_1 and w_2 and any reasons-apt tuples T_1 in w_1 and T_2 in w_2, if T_1 and T_2 are intrinsically and extrinsically *descriptively* indiscernible and w_1 and w_2 are *descriptively* indiscernible, then T_1 instantiates the favoring relation if and only if T_2 instantiates the favoring relation.

It may not seem like much of a change from Additive Reasons Supervenience, but it makes all the difference for the irreduciblist's avoidance of Brute Entailment.

In particular, the irreduciblist can avoid Brute Entailment if she accepts some version of the following claim:

> *Thick Essence of Reasons*: There is a descriptive way W of relating to a thick property such that what it is for a reasons-apt tuple T to instantiate the favoring relation is for there to be a thick property P such that T relates to P in the W way.

Thick Essence of Reasons is confusing. But all it says is this. Johnnie's forgetting to do the dishes is a reason for me to have an uncomfortable talk with him, but it's not a reason to stab him with a kitchen knife. What's the difference? Thick Essence of Reasons says that the difference has to do with a difference in the way that the distinct pairs of facts and actions relate to thick concepts.[7] Johnnie's forgetting to do the dishes is a reason for me to have an uncomfortable talk with him because there's at least one thick property that's relevantly descriptively related to that fact-action pair; and Johnnie's forgetting to do the dishes is not a reason for me to stab him with the knife because there is no thick property that's relevantly descriptively related to that fact-action pair.

[7] I leave out the agents and circumstances for the sake of easier comprehension.

If Thick Essence of Reasons is true, then TRS provides us with the resources to meet all of the four desiderata I listed in §2.6.1 that we need supervenience in order to satisfy. Those four desiderata are: supervenience is intuitive; supervenience is a conceptual constraint on the distribution of normative properties; supervenience is needed for normative explanation; supervenience is required to make sure that it's not arbitrary that normative reality is the way that it is.

First of all, we can satisfy the intuitions that two tyrannical dictators are both morally bad by appealing to a similarity in their thick properties—e.g., that both are cruel and did cruel things.

Second, Thick Essence of Reasons is, I will argue, a conceptual truth. Thick Essence of Reasons also entails TRS. Since it also entails TRS, TRS also comes out as a conceptual truth. Thick Essence of Reasons entails TRS because Thick Essence of Reasons tells us what reasons are: they're facts that are part of reasons-apt tuples that bear a specific relation to a thick property. Any time you have a reasons-apt tuple that bears that specific relation to a thick property, the fact in that reasons-apt tuple has to be a reason. That's what it is to be a reason. Now replicate all the non-normative properties and the thick properties and the descriptive relations between them. In any such world, the counterparts to the original reasons-apt tuple, thick property, and specific descriptive relation between them will all be there exactly the same as they were. So the counterpart to the reasons-apt tuple has to give us a reason as well. But that's just to say that Thick Essence of Reasons entails TRS. So Thick Essence of Reasons entails TRS, and TRS is therefore a conceptual truth given that Thick Essence of Reasons is a conceptual truth.

Third, Since TRS is a conceptual truth, we can fully explain why any particular reasons-

apt tuple is a reason by appealing to TRS and to the distribution of the descriptive properties. Why is such-and-such fact a reason? Well, it is part of a reasons-apt tuple that bears the relevant relation to a thick property. It's a conceptual truth that any such fact is a reason. So such-and-such fact is a reason. That's a full and completely adequate explanation.

Fourth, the truth of TRS fixes normative reality to descriptive reality so that, for example, it is not possible for there to be a world descriptively just like ours except that act utilitarianism is true. Furthermore, if Thick Essence of Reasons is true (and we can explain its truth), then we have an explanation of why the descriptive nature of the world fixes the normative nature of the world with metaphysical necessity. There are two kinds of normative properties: thick properties and thin properties. I'll start with thick properties. Thick properties *are* descriptive properties, so obviously any complete fixing of the descriptive properties will also fix the thick properties. And they will do so with metaphysical necessity because it's metaphysically impossible (and also incoherent) for all the descriptive features of the world to be one way and also a different way. If you like the metaphors we used to describe metaphysical necessity in §3.1, the thick properties come for free with the descriptive properties. Once God makes all the descriptive property instantiations, there's nothing more he has to do to make thick property instantiations. He's already done that. Nothing more to do.

Now take the thin properties. These include the favoring relation as well as the properties of being good, being right, and so on. We'll start with the favoring relation. Once you have all of a world's descriptive nature, you'll have will have all the non-normative features, all the thick features, and all the descriptive ways that they relate to each other. If Thick Essence of Reasons is true (and we can explain its truth), for a fact to favor an action

just is for that fact and action to be relevantly descriptively related to a thick property. So once you have a world's descriptive nature, you'll also have all the descriptive ways that facts, actions and thick properties relate to each other. In other words, you have the favoring relation, since the favoring relation just is one of these descriptive ways that facts, actions and thick properties relate to each other. Once we have the distribution of the favoring relation, Reasons Thin Centrality secures the other thin normative properties as well.[8]

To summarize. I have just emptied a gift bag of claims into this section and argued that, if these claims were true, then the irreduciblist can give an explanation of why her supervenience claim expresses a metaphysical necessity. I had argued in §3.1 that a theory's not being able to give such an explanation would constitute strong reasons to reject that theory. I then argued that we had strong reasons to reject reasons fundamentalism because reasons fundamentalism, as far as I could tell, could not offer a compelling explanation of why its supervenience claim expressed a metaphysical necessity. The argument I have just given in this section therefore purports to show that, although brute entailment concerns are serious concerns for irreduciblists of the reasons fundamentalist variety, they need not be serious concerns for other irreduciblists who reject reasons fundamentalism. In particular, they need not be serious concerns for irreduciblists who accept some version of Thick Essence of Reasons.

I have not yet done two important things, however. First, I have not talked at all about the relation between the thick properties and purely descriptive (non-normative) properties

[8]Reasons Thin Centrality may not secure them with metaphysical necessity, but they would do so if we adopted a buck-passing account of the thin properties. If we adopt such an account, what it is for any thin property P is to be just a purely formal higher-order property of having properties that provide reasons of a certain kind. Once you have all the reasons, the higher-order properties of these reasons come for free; therefore, the thin properties would come for free as well.

that I've made such a big deal about. Second, I've given no reason to accept any of these claims except for the fact that it would be convenient for the irreduciblist if they were true. The audience that would take that to be a compelling reason to accept any of these claims is very small. Let's deal with these concerns one at a time. I will consider the relation between thick properties and purely descriptive (non-normative) properties in §5.3, and will give independent reasons for accepting my own particular version of Thick Essence of Reasons in §5.4 and §5.5. I'll then end in §5.6 with a discussion of how Reasons Thin Centrality fits into the sketch of a theory of normativity that I provide in §5.4 and §5.5.

5.3 Does Additive Supervenience entail Transformative Supervenience?

A problem for my view

Here is a supervenience claim that, if true, would be really bad for me:

> *Supervenience of Thick Properties*: For any possible worlds w_1 and w_2 and any individual i_1 in w_1 and i_2 in w_2, if i_1 and i_2 are intrinsically and extrinsically *non-normatively* indiscernible and w_1 and w_2 are *non-normatively* indiscernible, then, for any thick property P, i_1 instantiates P if and only if i_2 instantiates P.

Supervenience of Thick Properties says that the non-normative features of the world fix the thick normative features of the world. If this were true, then Transformative Reasons Supervenience would not be more permissive than Additive Supervenience after all. I had argued earlier that irreduciblists faced serious difficulties in explaining why Additive Su-

pervenience expressed a metaphysical necessity, and appealed to Transformative Reasons Supervenience and Thick Essence of Reasons in order to help the irreduciblist avoid this worry. But if Supervenience of Thick Properties is true, then my commitment to Transformative Reasons Supervenience would also entail a commitment to Additive Reasons Supervenience.

Why? Because supervenience is reflexive and transitive. Supervenience is reflexive: sets of properties always supervene on themselves. Supervenience is also transitive: if A fixes $A \wedge B$, and $A \wedge B$ fixes C, then A fixes C. Let A be the non-normative properties, B be the thick properties, and C be the favoring relation. If the non-normative properties fix the thick properties, as Supervenience of Thick Properties says they do, the non-normative properties fix both the thick properties and the non-normative properties (by reflexivity). And if, additionally, the non-normative properties and thick properties jointly fix the distribution of the favoring relation, as Transformative Reasons Supervenience says they do, then the non-normative properties will also fix the distribution of the favoring relation.

I, qua irreduciblist, would then run into the same problems with explaining why Additive Supervenience expresses a metaphysical necessity that I tried to avoid by adopting Transformative Supervenience. The particular question that I would not be able to answer is: what is it about the distribution of non-normative properties that make it the case that they fix the distribution of the favoring relation with metaphysical necessity? We could no longer appeal to the distribution of thick properties, since we would then still face the further question of what it is about the distribution of the non-normative properties that make it the case that they fix the distribution of thick properties with metaphysical necessity.

There is an intuitive case for Supervenience of Thick Properties. Here is a completely

non-normative description of Sarah. Sarah sits outside the campus steps at 12:24pm as part of a line of ten students carrying signs that described some of the economic failures of state support for capitalist modes of production. Sarah sees the policeman coming with the pepper spray. Sarah tries to hide her face with the lining of the turtleneck collar but does not move. The policeman walks up and down the line spraying all the protesters with pepper spray.[9] Surely Sarah is courageous. And surely anyone just in Sarah's position would also be courageous. As Zangwill would say, it's insane and irresponsible to think otherwise. That's the intuitive case.

In response, I argue that Supervenience of Thick Properties is either incoherent or false. As mentioned in §3.2.1, I agree with Bilgrami that we cannot coherently imagine entailment relations between normative and non-normative properties. Supervenience of Thick Properties states an entailment relation between non-normative properties and those normative properties that are also thick properties. So I think that Supervenience of Thick Properties is incoherent. However, I will not give an argument for my views that this is incoherent, both because Bilgrami already gives these arguments in (Bilgrami 2006) and also because I want to also convince philosopher's who don't accept those arguments that they should not accept Supervenience of Thick Properties either. So in this section, I will argue that Supervenience of Thick Properties is false.

My argument will adapt Merricks' argument against the supervenience of mental states on microphysical states to the case of thick properties. The general thrust of the argument is that many commonly accepted supervenience relations are false, and that their falseness is surprising but unproblematic. There is nothing about the falsity of Supervenience of

[9] I'm referring of course to the incident at UC Davis that occurred in 2011 (Today 2011).

Thick Properties that produces deleterious consequences for the irreduciblist so long as the irreduciblist accepts Transformative Reasons Supervenience.

Merricks' argument against microphysical supervenience

Supervenience is not wholly uncontroversial even in the less controversial cases, like the supervenience of table facts on facts about parts arranged tablewise. When I describe supervenience to undergraduates (and I don't think I am alone in this), I find it helpful to say something like the following. "Look at this thing. It's a table! And all of its parts, all the pieces that make it up, those are parts of a table. Now take all of those parts and make copies of them that are just like the originals, and put them all together in just the same way that these parts are put together now. What is that thing I just made?" My students have near exceptionlessly said that it is, in fact, a table. But perhaps this surface appearance is misleading. Take, for instance, the following modification of an argument by Trenton Merricks (Merricks 1998).

It is intuitively plausible that the property of being a table supervenes on the properties and relations of the table's basic parts.[10] So take some table, which we can call *Big Table*. If you make duplicates of all the basic parts of Big Table, preserving these parts' intrinsic properties and restricted part-to-part relations,[11] then the result, whatever else might be true of it, will also be a table. That's the intuition, at least, that all of my students have consistently had.

[10] We can construe these parts as whatever the fundamental material constituents of the universe happen to be.

[11] I am assuming, as Merricks does, that we share an intuitive notion of intrinsicality for the purposes of this argument, modulo all the appropriate and important caveats about the intrinsicality debate in contemporary metaphysics.

Now, annihilate one of the basic parts of Big Table. Intuitively, the addition or subtraction of a basic part makes no difference as to whether something is a table. So the parts that survive the annihilation of a single part will still compose a table, which we will call *Little Table*. Intuitively, again, the property of being a table supervenes on the intrinsic properties and relations of the table's parts. If you make a duplicate of all the basic parts of Little Table, preserving these parts' intrinsic properties and restricted part-to-part relations, then the result, whatever else might be true of it, will also be a table.

Little Table was generated by annihilating one part from Big Table and taking everything that was left over. We can further suppose that stripping away a single part from Big Table does not disrupt the other restricted part-to-part relations that obtain between the parts that remain. It is merely a single part, after all.[12] So Little Table is a proper part of Big Table. Now if it's true that the property of being a table supervenes on the intrinsic properties and restricted part-to-part relations of the parts that compose the table, then a duplicate of Little Table will be a table no matter what else might be true of the extrinsic situation. Well, one possible extrinsic situation is the situation in which Little Table constitutes a proper part of Big Table. The supervenience claim therefore entails that, if something is a duplicate of Big Table, not only is it itself a table, but it also has a proper part, a duplicate of Little Table, which is also a table. The supervenience claim therefore entails either that (a) if Big Table exists, then two tables exist;[13] or (b) there are no tables.[14] Alternatively, we can save

[12]This is, granted, an empirical question to which I don't know the answer. But even if the relations of the remaining parts would not be perfectly preserved, even for an instant, we can probably come up with some situation that would guarantee the preservation: Merlin's spell momentarily suspends certain alterations, say (Schaffer 2000), (Lewis 2000). This is a thought experiment.

[13]In fact, indefinitely many tables would exist, given the vast number of parts that we could choose to annihilate in the thought experiment

[14]The elimination of tables from our ontology would satisfy the supervenience claim, since any two du-

the common sense view that there is only one table by rejecting the supervenience claim.

Merrick's argument can be adapted and applied to the case of normative properties with the help of some assumptions about the metaphysics of action provided by Jennifer Hornsby in her very insightful article (Hornsby 2012). In my case, I am adapting Merrick's arguments against supervenience to the case of thick properties.[15] So the property that I use for this purpose needs to be a thick property. We'll take Sarah's courageous act of raising the turtleneck collar to her face.

Hornsby's metaphysics of action

Hornsby distinguishes actions into two types: activities and performances. Activities are "temporal stuff" in the sense that, for any activity, you can always take a proper temporal part of that activity and it will turn out that that proper part is still the same kind of activity as the larger activity of which it is a proper part. One example of an activity is walking. You're walking aimlessly around campus between 4am and 5am to clear your head after a night of dissertation writing. If you just take the part of that activity that extends from 4am to 4:30am, that proper part that you just isolated still qualifies as walking. Other examples of activities are painting, raising your arm, and so on.

Performances are further distinguished into two types: accomplishments and achievements. The difference between accomplishments and achievements has to do with dura-

plicates of any set of parts would be alike in their property of failing to compose a table.

[15] It came to my attention after finishing the first draft of this section that a similar application of Merricks' results had already been given in (Walsh 2011). However, Walsh aims only to argue that there are no intrinsic normative properties. My argument goes further: it aims to show that the distribution of thick properties does not supervene on the distribution of non-normative properties. Furthermore, my own argument is much more detailed and makes use of resources in the philosophy of action that Walsh does not use.

tion: accomplishments have a temporal duration, while achievements do not. Looking for a book is an accomplishment because it is a process that extends over time: you look under the covers, you look in the refrigerator (where, in my absent-mindedness and exhaustion from dissertation-writing, I frequently leave many things) and so on. *Finding* the book is an achievement: it is something that occurs in an instant. Similarly, running a 100-meter dash is an accomplishment, and making it across that 100-meter dash is an achievement. Achievements are time-slices of accomplishments that culminate and complete the accomplishments.

Accomplishments are sortal events that are composed of activity stuff. My accomplishment of running the race is composed of the activity of running. Accomplishments are distinct from mere stretches of activity in that any accomplishment, unlike any activity, involves some kind of outcome such that, if that outcome occurs, the outcome will be an achievement of the accomplishment and mark a completion and culmination of that accomplishment. Stretches of activity do not involve any outcome that marks their completion. To see the difference, imagine that you're running on the racetrack but the referee took away the finish line, so that there is no longer any way to win the race. Your running the 100 meters would still be a stretch of running activity. But it would no longer be an accomplishment because there is no longer any outcome such that achieving that outcome would mark a culmination to what you're doing. Another way to see the difference is to imagine continuing to run past the finish line. You're still running after you pass the finish line, but any running you do past the finish line will not be part of the accomplishment of running the 100-meter race. Running the 100-meter race was over when you crossed the finish line,

even though your running was not over.[16]

Accomplishments can be incomplete. I can look for my book and give up before finding it, and I can also be running a 100-meter dash and get vaporized by a nuclear weapon before making it across the finish line. In these cases, the accomplishment is the stretch of activity that begins with the beginning of your engaging in the activity, and ends when the stretch of activity ends.

Sometimes there are many time-slices that satisfy the achievement conditions of an accomplishment. If Aldous Huxley stopped writing *Brave New World* one word sooner than he did, his accomplishment of writing *Brave New World* would have culminated in the achievement of *Brave New World*'s having been written.[17] However, since Huxley continued to write anyway, the actual time-slice that constitutes the achievement of having written *Brave New World* would be the instant that Huxley lifts up the pen for the last time. After all, right before he lifts up the pen, Huxley's still writing *Brave New World*. Although Huxley could have accomplished the writing of *Brave New World* sooner simply by lifting his pen, the fact that he continued to write anyway means that, at the time that he is writing that last period, *Brave New World* is still being written. It's not done yet. I imagine that any academic reading this would recognize the feeling.[18]

Importantly, accomplishments are not stuff. They are not such that any proper part of an accomplishment is itself the same accomplishment. When you are running the 100-meter

[16]I think that using "accomplishment" to denote the kind of action that it does is confusing. I use the term "accomplishment" because that's the term that Hornsby uses. Hornsby herself takes it from (Mourelatos 1978).

[17]After looking at the last few words of the book again, I think that this is a safe assumption to make.

[18]This paragraph represents my own argument. As far as I know, Hornsby does not herself make this point.

dash, there are many stretches of running activity, but there is only one instance of running a 100-meter dash.

With that background in place, we are now in a position to construct our argument against the supervenience of thick properties on non-normative properties.

Why thick properties don't supervene on non-normative properties

Sarah's action of raising her turtleneck collar is an accomplishment that is composed of the activity of raising her collar and culminates with the collar's being raised. Call this accomplishment *Big Accomplishment*.

We all agree that something about what Sarah did is courageous. The options for what about her action is courageous are: the achievement of the raised collar, the accomplishment of raising the collar to her face, or the activity of collar raising that constitutes the accomplishment. The courageous thing in this case happens to be the accomplishment (Big Accomplishment). The achievement itself is not courageous, since what Sarah did would still be courageous if suddenly her arm started to shake from fear and she didn't manage to raise the collar all the way. Furthermore, there are two reasons why we should not think that the activity itself is courageous. Activity is infinitely divisible, so Sarah is engaging in an infinite number of stretches of collar-raising activity. If what was courageous was the activity, Sarah would therefore be doing in fact an infinite number of courageous things. What Sarah did was very courageous, but there is still only one courageous thing that Sarah did. Secondly, suppose Sarah pressed the collar to her face even harder than she did. In this case, there is more collar raising activity stuff. However, it is implausible to think that

Sarah would be more courageous for pressing the collar to her face harder. She is already courageous enough as she is. And she would not be less courageous if she engaged in less collar-raising activity because her arm started shaking from fear.

Now imagine a world that's exactly the same as Sarah's except that, in this other world, Sarah's counterpart does not press the collar as hard against her face as she did. Call this world *World 1*. Let Sarah's world (not the world of Sarah's counterpart, but the world of Sarah) be called *Big World*. In World 1, Sarah's counterpart is still doing an action that's an accomplishment, since her collar-raising is oriented towards the outcome of having a raised collar. Sarah's counterpart's accomplishment is also complete, since it culminates with the achievement of having the collar raised to her face. Call this accomplishment in World 1 *Little World 1 Accomplishment*. Furthermore, her accomplishment is still courageous, since, intuitively, the loss of a little bit of collar-raising activity would not change the fact that what Sarah did is courageous.

Suppose furthermore that the only thing that's different about Big World and World 1 is this difference in the presence or absence of the little bit of collar-raising activity. In other words, nothing about the absence of a small bit of collar-raising did anything to make anything else about World 1 any different from the way that its counterparts are in Big World.[19] Given this, all of the activity that constitutes Little Accomplishment is intrinsically just the same as its counterpart in Big Accomplishment. In other words, there is a proper

[19]Suppose there's a wizard or what have you keeping the lack of this activity from having any other effect on anything else. The actual biology and physics of what's going on does not matter because this argument and Merricks' argument are both arguments that proceed from thought experiment. I also get to talk about wizards having strange causal powers because Jonathan Schaffer gets to (Schaffer 2000), even in *The Journal of Philosophy*; and, since he is a much better philosopher than I am, it would be unfair to give him access to more philosophical tools than me. Whether this qualifies as a joke will depend on whether it's an essential property of jokes that they be funny. I just feel compelled to attempt a joke because the example I chose for my argument is so depressing.

part of Big Accomplishment that is an intrinsic duplicate of Little Accomplishment. Call this proper part of Big Accomplishment *Big Accomplishment*$^{-}$.

Now that we have all this in place, we need one further assumption, one that Lewis defends in (Lewis 1986, pp. 86-90). That assumption is that, for any object or set of objects in a world, there is another world that contains just intrinsic duplicates of that object or set of objects. For example, there is a world that contains only me and my intrinsic properties; there is a world containing only me and my intrinsic properties plus this computer and all its intrinsic properties, and so on.[20]

Take all of Big World, minus the bit of collar-raising activity that didn't have a counterpart in the world with Little Accomplishment. In other words, you'll take everything about Big World, take out Big Accomplishment, and then include Big Accomplishment^{-}. With all that stuff you now have, make an intrinsic duplicate of it. Call that intrinsic duplicate *World 2*, and let *Little World 2 Accomplishment* be the counterpart in World 2 of Big Accomplishment^{-}. There is such a world so long as we accept Lewis' duplication principle.

If we compare World 2 (the world with Little World 1 Accomplishment) to World 1 (the world with Little World 1 Accomplishment), they are very much the same. In both worlds, there is a stretch of collar-raising activity, and in both worlds these respective stretches of collar-raising activity constitute an accomplishment that ends with the achievement of a raised collar.

There is, however, one important difference between World 1 and World 2. Little World 1 Accomplishment is courageous. However, Little World 2 Accomplishment is not. Why? Because Little World 2 Accomplishment is stipulated to be an intrinsic duplicate of Big

[20] I get this idea from Merricks.

Accomplishment^{-}. However, Big Accomplishment^{-} is a proper part of Big Accomplishment, and I earlier argued that the proper parts of courageous accomplishments are not themselves courageous.

In other words, the only difference between World 1 and World 2 is that Little World 1 Accomplishment is courageous and Little World 2 Accomplishment is not. The only difference is difference in the distribution of thick properties. There are no non-normative differences between World 1 and World 2.

Here is Supervenience of Thick Properties:

> *Supervenience of Thick Properties*: For any possible worlds w_1 and w_2 and any individual i_1 in w_1 and i_2 in w_2, if i_1 and i_2 are intrinsically and extrinsically *non-normatively* indiscernible and w_1 and w_2 are *non-normatively* indiscernible, then, for any thick property P, i_1 instantiates P if and only if i_2 instantiates P.

Here is the result after we substitute in our examples:

> *Surprise!*: If Little World 1 Accomplishment and Little World 2 Accomplishment are intrinsically and extrinsically non-normatively indiscernible and World 1 and World 2 are non-normatively indiscernible, then Little World 1 Accomplishment is courageous if and only if Little World 2 Accomplishment is courageous.

However, I just showed that, given Hornsby's background philosophy of action and one assumption of David Lewis, Merricks' argument against microphysical supervenience can by extended to show that there are no non-normative differences between World 1 and World 2, and that Little World 1 Accomplishment is courageous while Little World 2 Accomplishment is not. Therefore, on these assumptions, the antecedent of Surprise! is true

while the consequent is false. So Surprise! itself is false, giving us a counterexample to Supervenience of Thick Properties.[21]

Thick property supervenience failure is not bad

One might object that what I have given is an abstract argument and therefore inappropriate in the contexts of arguments about normative phenomena. I am very sympathetic to the thought that certain normative phenomena are not well captured by overly theoretical characterizations. This is an important insight of Williams, Stocker, Gilligan, and other figures in the moral anti-theory tradition. However, normative supervenience is itself an abstract doctrine, and as such, abstract arguments against it seem perfectly appropriate.[22]

Note that the Merricks style of argument can also be used by the reasons fundamentalist to deny Additive Reasons Supervenience. However, denying Additive Reasons Supervenience is problematic for the reasons fundamentalist because the reasons fundamentalist is committed to Additive Reasons Supervenience for the reason that he has no other supervenience relation that he can rely on to satisfy the desiderata I described in §2.6.1 that we need

[21] Opponents of Merrick's argument against microphysical supervenience sometimes argue that properties such as being a table are *maximal* properties. A maximal property P is a property such that, for any fusion F that satisfies P's other instantiation conditions, there is no fusion of which F is a proper part and that also satisfies P's other instantiation conditions (Sider 2001), (Sider 2003). Maximal properties are extrinsic properties because whether or not a fusion F exemplifies them depends on the presence or absence of other things outside F. Sider raises the issue of maximality in order to object to Merricks' argument against microphysical supervenience. Merricks responds in (Merricks 2003). I will not say more about this debate other than to remark that Merricks seems to me to have the better argument.
Also note that the argument does not go through if one rejects Hornsby's categorization of actions or one rejects Lewis' assumption that, for any object or set of objects, there is a world containing nothing except an intrinsic duplicate of this object or set of objects. I cannot argue for these claims here. I will just say that I and many other people think that these claims are correct and that there are good reasons for thinking so, which I leave to Hornsby and Lewis to defend far more adroitly than I could.

[22] Furthermore, anyone who thought that abstract arguments were inappropriate for discussions of normative phenomena would have stopped reading this dissertation a long time ago.

supervenience to be able to satisfy. I avoid this problem because I accept Transformative Reasons Supervenience.

Further, the argument I just gave against Supervenience of Thick Properties does nothing to show that Transformative Reasons Supervenience is false. Although, as I argued, World 1 and World 2 are non-normatively indiscernible, they are not *descriptively* indiscernible, since they differ in that Little World 1 Accomplishment exemplifies the thick (and therefore descriptive) property of being courageous, while Little World 2 Accomplishment does not exemplify the thick (and therefore descriptive) property of being courageous.

The falsity of Supervenience of Thick Properties is surely an odd result. However, to the extent that its oddness causes problems, these problems generalize outside the normative case. They're not problems that are unique to normativity, and certainly not problems that are unique to my own view. Merricks' argument can be used to show the failure of supervenience in a wide variety of cases, such as cases of ordinary objects and the case of consciousness. Anyone who wants to defend the existence of ordinary objects at all has to deal with this result, and one way to deal with this result is to give up supervenience. I think the moral that we should draw from these results is not that failures of supervenience generate problems, but that philosophers generate problems when they place too much importance on supervenience. I think that supervenience does not have the kind of wide application that many philosophers take it to have, and that its importance should be downplayed unless there are compelling reasons for it. In the normative case, there are compelling reasons for accepting some sort of supervenience claim: it's a conceptual constraint, it's presupposed by our practices of normative explanation, and it's needed to keep

the normative world from being "out of control" (as Shafer-Landau says).[23] None of this is true, for example, in the case of the strong supervenience of consciousness on microphysical properties.[24]

There may be strong reasons to accept the supervenience of tables on microphysical properties. Merricks argues that, if there are tables, this supervenience claim generates inconsistencies. Merricks, however, thinks that there are not strong reasons to accept the existence of tables, and so he eliminates them from his ontology. I don't think that we can take the analogously drastic step of eliminating normativity.[25] So if there were a tension between normativity and supervenience, I think that we should keep normativity and give up supervenience. This is what I am advocating in the case of the supervenience of thick properties on non-normative properties. In doing this, I don't take myself to be in any worse of a situation than anyone who wants to defend the existence of ordinary objects despite the tension between their existence and supervenience. Thus the problem, if it is a problem, is not a unique problem of my view. Fortunately, we don't need to reject normative supervenience entirely, since we can rely on Transformative Reasons Supervenience to do the theoretical work that we need supervenience to do.

In addition to the previous "partners in guilt" strategy for explaining why denying the supervenience of thick properties on non-normative properties is not a problem, there is

[23]This is good for me, since otherwise the hundred pages of this dissertation that discuss normative supervenience would have been a complete waste of time. I discussed these compelling reasons to accept supervenience earlier in §2.6.1.

[24]In brief, phenomenal and microphysical properties seem far too different and conceptually autonomous for there to be equally compelling reasons to accept supervenience in the case of consciousness as there are reasons to accept supervenience in the normative case. For why, read (Chalmers. 1996) or chapters 2, 3 and 6 of Descartes' *Meditations*.

[25]Obviously plenty of people don't agree with this, but I think that they are wrong.

at least one more reason to think that the failure of supervenience of thick properties on non-normative properties is not a problem. Thick properties have both a descriptive and normative nature. Suppose that you could find a set of descriptive properties D that gives you the entire, exhaustive descriptive content of courage. If you had such a property, you are guaranteed to be able to isolate a thin normative property N that gives you the entire normative content of courage. N would be whatever is left over if you start with courage and take away D from it. N is a thin normative property because, after all the descriptive content is removed, the only content that remains is normative. Perhaps N is just the property of being good; perhaps it is some exotic thin property that only courageous things instantiate. It doesn't matter; the point is that N is a thin normative property. Call this a *conjunctive* view of the thick property of courage. On the conjunctive view of courage, courage is identical to a conjunction of a thin normative property and a purely descriptive (non-normative) property. Courage is identical to $D \wedge N$, where N is whatever is left over once the descriptive nature of courage is stripped away.

If the conjunctive view of courage is correct, then it does seem problematic to deny Supervenience of Thick Properties. World 1 and World 2 from the Sarah example above were shown to be non-normatively indiscernible. The only difference between them was a difference in their thick properties. On the conjunctive theory of courage, however, the descriptive content D of courage is also non-normative. So, since World 1 and World 2 are non-normatively indiscernible, D must be present in both World 1 and World 2, leaving the difference in the thin normative property N as the only difference between the worlds. There are, in other words, no descriptive differences, but only differences in the presence or absence of thin normative properties. Failure of Supervenience of Thick Properties would

therefore amount to a failure of Additive Reasons Supervenience. In that case, we are in the same position as the reasons fundamentalist who needs Additive Reasons Supervenience to avoid all the problematical consequences of rejecting supervenience that I discussed in §2.6.1.

The conjunctive view of thick properties is not the only possible view of thick properties. We can deny that there exists any purely descriptive (non-normative) property D that gives you the entire exhaustive descriptive content of courage. Here is what such a picture of courage would look like. All the platitudes about courage are true: that courage is exemplified only when you are afraid, that William Wallace is courageous, that enduring pain that you can avoid in order to achieve some great feat is courageous, and so on. These are non-normative statements that tell us something about what courage is non-normatively like. But these don't tell you the whole story. There is something else about courageous acts that all these platitudes put together won't capture: that courageous acts are *courageous*. Put together any purely descriptive property D you like, and I will be able to find an example of something that's courageous but not D, or D but not courageous.

Another way to think about this is to compare it to Williamson's strategy in *Knowledge and its Limits* (Williamson 2002). In that book, Williamson argues that the infamous Gettier cases and the failure of all the literature that ensued to come up with any definition of knowledge that avoided all Gettier-style counterexamples shows that "knowledge" is unanalyzable. However, "knowledge" is a key concept in epistemology that does theoretical work that other epistemological concepts do not do, so we can't simply eliminate knowledge and remove it from our theory of the furniture of the mind. The indispensibility of our concept of knowledge to epistemology gives us a reason to avoid the eliminativist route

and accept that knowledge is a genuine mental state.

I am suggesting that something similar may be true of some thick concepts: namely, that we can't give an analysis of them as a conjunction of purely descriptive and thin normative concepts. However, thick concepts are key concepts in normative ethics that do theoretical work that thin normative concepts do not and cannot do, so we can't simply avoid the worries about a lack of analysis by eliminating thick concepts from our theory of normative reality. The indispensibility of thick concepts to normative ethics gives us a reason to avoid the eliminativist route and accept that thick properties are genuine features of normative reality.[26]

Given this, the denial of Supervenience of Thick Properties seems significantly less problematic. There is no problem in general with imagining two scenarios that differ in that one possesses a descriptive property that the other lacks. For instance, General Relativity allows the stress-energy tensor to be null at regions. Imagine two regions that are exactly alike except that one has a null stress-energy tensor value and another lacks the property of having a null stress-energy tensor value. There is no problem here. Similarly, take two leptons that are exactly alike except that one has the property of having a positive charge and the other lacks this property. Again, there is no problem here.

These examples bear one important disanalogy to the conclusion I wanted to draw from the Merricks-style argument I gave about Sarah's courage. Two regions that differ with respect only to their having or lacking a null stress tensor value, and two leptons that differ only with respect to their having or lacking a positive charge, will nevertheless bear different

[26]These remarks do not constitute an argument for the position I am describing. They are merely sketching a position in logical space that you can take towards thick properties, one that I am sympathetic with but cannot develop here because the dissertation is already too long.

causal relations to other things. Null stress-energy tensor regions will produce different gravitational effects on other regions of the universe, and positively-charged leptons will interact differently with other particles than their negatively-charged counterparts. Thus, two worlds that differ with respect to the presence or absence of these properties in a single set of counterpart regions or a single set of counterpart leptons will also differ with respect to many other non-normative features. What I aimed to establish earlier is that there are two worlds that are non-normatively discernible but discernible with respect to whether Sarah's accomplishment is courageous. Courage, as a descriptive property, either causally relates to other things in the universe, or it does not causally relate to other things in the universe. If it does causally relate to other things in the universe, then the presence or absence of courage between two worlds seems to entail that there will be changes in the overall non-normative nature of these two worlds. The two worlds then would not constitute a counterexample to Supervenience of Thick Properties. If, however, courage does not causally relate to other things in the universe, then we face an epistemic problem similar to the one described in §2.5.1. There, I mentioned that not even Moore thinks that we have a faculty of moral intuition robust enough to determine whether two non-normatively indiscernible things differ with respect to their normative properties. It is difficult to see how we would know whether we were in the world in which Sarah's act is courageous or the world in which her act is not courageous. This is important because, on my view, thick properties constitute reasons. So, if we don't know whether a particular thick property is present or absent, then we can't know exactly what reasons we have to act in particular ways.

A more analogous example to the courage case I've been discussing would be two

worlds that differed with respect to the presence or absence of phenomenal properties. Chalmers' zombie world, which I discussed in §4, is non-phenomenally indiscernible from ours and differs from ours only in that our world features phenomenal property instantiations and the zombie world does not. Furthermore, the property of being blue arguably does bear causal relations to other aspects of the universe. You might think, for example, that a reliable subject perceives blue objects as blue because *these objects are blue*: it's their blueness that causes the perceiver's perception of the object to be a blue experience. The perceiver's perception of blue is also a phenomenal property, so, in this case, the phenomenal property of blueness would bear causal relationships to other phenomenal properties such as blueness-perception. If this were true, blueness can alter the causal profile of a world without altering the non-phenomenal causal profile of a world.

Normative properties may function in a similar way. Take Johnnie, who sees Sarah's courage and walks up to the line to sit next to her. Frequently, we explain our and other people's actions by citing the normative reasons that they have. We could also of course cite features of Johnnie's psychology—we could say, for instance, that Johnnie desired to overthrow capitalism and believed that subjecting himself to pepper spray on this occasion would help him attain that goal (because the pepper spraying would occur in the context of an anti-capitalistic protest, violence directed at people in anti-capitalist protests has such-and-such beneficial effects on the status of the ongoing revolution to overthrow capitalism, and so on). We could also cite features of his neurophysiology: Johnnie's sensory modalities interacted with the gross motor cortex in such-and-such ways, and so on. But there is another kind of sui generis, distinctly normative explanation of action that purports to explain the occurrence of actions by citing features of normative reality. Why did Johnnie

join the protesters on the line? Because Sarah's courage gave him reasons to do so. He joined the line for those reasons.

Furthermore, Johnnie's joining the line was a courageous act. Since Johnnie joined the line because of Sarah's courage and the reasons Sarah's courage provided, there's a sense in which Sarah's courage bears causal relations to other essentially normative events such as other courageous acts. In other words, Johnnie performed a courageous act because of Sarah's courage. If Sarah were not courageous (if, say, Sarah and Johnnie were in a world that did not feature thick property exemplifications), then Johnny would not have performed a courageous act. He may have still joined the line, but his joining the line would not be courageous.

I mentioned earlier that there is a view according to which phenomenal properties only bear causal relations to other phenomenal properties and give rise only to phenomenal events. If we extend this view to the case of normative properties, we can say that thick properties do not bear causal relations to non-normative properties and give rise only to essentially normative events. There is therefore a position in logical space that allows for the following two claims to be true. First, there are two possible worlds that non-normatively indiscernible but discernible in that one world features thick property instantiations while the other does not. Second, in the world that features thick property instantiations, thick properties are causally isolated from non-normative events, but are nevertheless causally efficacious in the sense that they bring about *essentially normative* events. By "essentially normative events", I have in mind actions like Johnny's courageous act, but there may be other essentially normative events as well, such as the systematic playing out of injustice. Let *normative causation* be any instance of causation between two essentially normative

events.[27]

It is useful to compare this kind of causal picture to the picture of reality that Spinoza gives. For Spinoza, the world is identical to a single, monistic substance whose essence is able to be conceived in different ways. These different ways of conceiving of the essence of the monistic substance are called *attributes*. We can conceive of the monistic substance in an extended way, in which case we would conceive of events as causal interactions between essentially extended things: the corpuscles bouncing around and so on. We could also conceive of the monistic substance in a thinking way, in which case we would conceive of events as causal interactions between "essentially thinking things", whose nature Spinoza is not particularly clear about but which perhaps mean something like things that act as substrata for phenomenal properties.

Spinoza also accepts the doctrine of causal parallelism, according to which, for every causal interaction of the extension type between two modes of the monistic substances, there is also a causal interaction of the thinking type between those two modes of the monistic substance. Causal parallelism also entails that there are no hybrid causal interactions. In other words, for Spinoza, it is never the case that an extended thing can cause changes in thinking things, and it is never the case that a thinking thing can cause changes in extended things.

Spinoza also holds that there are an infinite number of ways of conceiving of the essence

[27]I do not have very worked out views on the nature of causation in general or normative causation in particular, but I suspect that the difference between normative causation and "non-normative" causation would not reside in the nature of the causation at play, but rather in the nature of the entities standing in the causal relations. In other words, what I tend to think is that there is a single notion of cause which would be at work in both normative and non-normative instances of causation, and that the difference between normative and non-normative causation lies simply in the fact that, in cases of normative causation, the events standing in causal relations are essentially normative, whereas in cases of non-normative causation, the events standing in causal relations are not essentially normative.

of the monistic substance but that we, with our finite understanding capacities, cannot conceive of the monistic substance in more than two ways (the thinking way and the extended way). There may, however, be a third way of conceiving of reality that Spinoza does not countenance: a normative way. In conceiving of the situation on the protest line, we can conceive of Sarah and Johnnie as essentially extended beings causally relating to each other according to the laws of extension (physics, neurophysiology and so on). We can also conceive of them as essentially thinking beings, causally relating to each other according to the laws of thought (i.e. the laws of folk psychology). Alternatively, we can conceive of them as essentially normative beings, causally relating to each other according to the laws of normativity.

The last feature of Spinoza's view that's relevant for my purposes here is that, for Spinoza, our finite understanding is such that we cannot conceive of the essence of the world through two distinct attributes at the same time. For Spinoza, we cannot conceive of Johnnie and Sarah's interaction as both an interaction between essentially extended things and, at the same time, as an interaction between essentially thinking things. If we allow that normativity constitutes a third attribute—that is, a third way of conceiving of the essence of the monistic substance—it would follow on Spinoza's view that we cannot conceive of Johnnie and Sarah's interaction both as an interaction between essentially extended things and, at the same time, as an interaction between essentially normative things. The kind of view that emerges would bear some similarities to Bilgrami's view, mentioned towards the end of §3, that relations between normative and non-normative properties are not coherent.[28]

[28]I call Transformative Reasons Supervenience *Transformative* Reasons Supervenience because of this

As I mentioned earlier, I accept that relations between normative and non-normative properties are not coherent. I am sympathetic to Spinoza's view that we cannot conceive of the essence of the world through two distinct attributes at the same time. If we cannot conceive of the essence of the world through two distinct attributes at the same time, then it will not be possible to conceive of our world—or any world—as having both normative properties and extended (non-normative) properties. Conceiving of Supervenience of Thick Properties requires that we conceive of worlds that have both thick (i.e. normative) properties and non-normative properties. Conceiving of my denial of Supervenience of Thick Properties requires conceiving of Supervenience of Thick Concepts and then running it through a negation operator. So, if Spinoza is right that we cannot conceive of the essence of the world through two distinct attributes at the same time, my denial of Supervenience of Thick Properties turns out to be incoherent.

However, we do not have to accept Spinoza's view that it is not possible to conceive of distinct attributes at the same time. Chalmers, for instance, would not, since he thinks that there is no problem in (positively ideally) conceiving of our world at a time as, for instance, such that its members have (a) non-phenomenal properties in virtue of which the members

aspect of Spinoza's view. For Spinoza, conceiving of the world first through the attribute of extension, and then through the attribute of thought, involves a kind of transformation in our way of conceiving of the world. Although in both cases, we are conceiving of the same world with the same underlying causal structure, the essential nature of this world and its causal structure is conceived to be radically different. Now suppose we allow, as I'm suggesting, that there be a third, essentially normative way of conceiving of the world. Conceiving of the world first through the attribute of extension and then through the attribute of normativity will involve a similar transformation in the way that we conceive of the world. Transformative Reasons Supervenience is "transformative" in the sense that it can be formulated and understood when one is conceiving of the world through the attribute of normativity. This is the case because Transformative Reasons Supervenience construes the supervenience relation as a relation between normative properties and descriptive properties. Descriptive properties are also normative properties in the case when the descriptive properties are thick properties. When we conceive of the world through the attribute of normativity, we can access all and only the normative features of the world and the relations between them. We can therefore understand what it would mean for the distribution of normative properties to supervene on the distribution of descriptive properties in the case in which the descriptive properties include only thick properties.

causally interrelate in a non-phenomenal way, and also have (b) phenomenal properties in virtue of which its members causally interrelate in a phenomenal way. Chalmers also thinks that there is no problem in conceiving of a world that is just like ours in non-phenomenal respects but lacks phenomenal features. Similarly, I am claiming here that we can conceive of our world as such that its members have (a) non-normative properties in virtue of which the members causally interrelate in a non-normative way, and also have (b) normative properties in virtue of which its members causally interrelate in a normative way. I also think that there is no problem in conceiving of a world that is just like ours in non-normative respects but lacks thick properties and therefore lacks any normative features at all. I also argued in §5.3.4 that these distinct conceptions are conceptions of distinct possible worlds and therefore that Supervenience of Thick Properties is false.

This section has attempted to show that denying Supervenience of Thick Properties is not unacceptable or fatal but actually points to an attractive sketch of a theory of normativity that we have good reasons to develop further. On the theory of normativity sketched here, thick properties, while causally isolated from non-normative properties and essentially non-normative events, stand in causal relations with each other and give rise to all that is normative.

A theory like this would be attractive for a variety of reasons. First, it would preserve the kind of autonomy of normative reality from non-normative reality that provides probably the most theoretical motivation for adopting irreduciblist views in general.[29] Second, it is consistent with a supervenience claim (Transformative Reasons Supervenience), which

[29]Scanlon's own domain-specific theory discussed in §2.4.2 also allows for the autonomy of normative theorizing, but it does not give any metaphysical weight to normative reality.

provides us with a way to satisfy the desiderata that I argued in §2.5 we needed a supervenience claim to satisfy. Third, it places thick properties at the center of normative reality. Why this is an advantage will be discussed further in §5.5.

I argued in §5.2 that irreduciblists can avoid worries about brute entailment by introducing thick properties into their theory of normative reality and arguing that reasons, essentially, are facts that bear some specified relation to thick properties. I did not yet give any account of what I take that relation to be. In the next sections §5.4 and §5.5, I'll argue for my own theory of what the relation is that reasons essentially bear to thick properties. The general upshot of the argument is that the essential relation is a species of the relation that holds between questions and answers.

5.4 Reasons as answers

I argue in this section that reasons are *answers*. After all, it seems like a natural place to start. If you notice the rain outside and are considering taking an umbrella, a reason like "it's raining outside" strikes us as an appropriate (although perhaps not complete or correct) response to the question why you should take the umbrella. Yet despite the obviousness with which this strikes us, there is surprisingly little discussion in the literature about what it is about reasons that explains their power to function as correct answers to why-questions.[30] The reasons-as-answers approach that I aim to defend gives a clear explanation of why this

[30]The only place that I know of to look for contemporary discussion of this thesis is in an exchange between Hieronymi and Schroeder at (Hieronymi 2005), (Schroeder 2010), and (Hieronymi 2013). Even this exchange places the reasons-as-answers thesis as secondary to the exhaustively discussed wrong kind of reason problem. Like most everyone else in philosophy, I also have many things to say about the wrong kind of reason question, but will not be discussing the wrong kind of reason problem in this dissertation.

is: on the view, reasons just are answers to certain kinds of why-questions.

The idea that reasons are answers to why-questions has a history: it is arguably implicitly present in a claim Anscombe makes about reasons. Says Anscombe:

> What distinguishes actions which are intentional from those which are not? The answer that I shall suggest is that they are the actions to which a certain sense of the question "Why?" is given application; the sense is of course that in which the answer, if positive, gives a reason for acting. But this is not a sufficient statement, because the question "What is the relevant sense of the question "Why?" and "What is meant by 'reason for acting'?" are one and the same. (Anscombe 1957, p. 9)

For Anscombe, an action ϕ is intentional just if the answer to the question "Why did you ϕ?" (when "you" is indexed to whomever did the ϕ-ing) gives a reason for acting. The claim seems to be not only that reasons for action are the sorts of things that can be given by answers to certain kinds of why-questions, but that "reason for action" is to be understood as meaning the same thing as "answer to a [certain kind of] why-question", after the relevant kind of why-question is sufficiently worked out.

Anscombe here is talking about motivating reasons: reasons why one *did* something or *would have done* something. Motivating reasons are not (or at least need not be in every case) normative reasons, which are reasons why one *ought to have done* something. My dissertation is exclusively discussing normative reasons and will continue to do so until the end. So, in this section, the reasons-as-answers account will be distinct from Anscombe's in that my account will be an account of normative reasons rather than of motivating reasons.

In the course of arguing for the view that reasons are answers, I criticize Scanlon's view that the reasons relation is a four-place relation. My argumentative strategy is to isolate particular situations such that the four-place view of the reasons relation generates

intuitive inconsistencies. Because the view generates inconsistencies, the view needs to be either rejected or modified. My particular approach is to modify the view by adding additional elements into our theory of the the logical structure of the reasons relation. In other words, I argue that the reasons relation is a six-place relation rather than a four-place relation, and that my particular characterization of the two additional components avoids the inconsistency worries that I raise against Scanlon's view.

I do think that, in the significant majority of cases, Scanlon's view that the reasons relation is a four-place relation would not generate the kinds of counterexamples that I give throughout the argument of this section. In other words, if the space of reasons were more constrained than it actually is, Scanlon's view that the reasons relation is a four-place relation would have been perfectly adequate. However, what I and Scanlon both seek is a general theory of the logical structure of reasons, one that can accommodate all cases. The view that I will give in the course of this section better qualifies as a general theory of the logical structure of reasons because, as I argue, it is able to accommodate strictly more cases than Scanlon's theory.

Question structures

In order to formulate and evaluate the prospects for an answers account of reasons, we first need a working account of what why questions are. The account of why questions that I will adopt is the one offered by van Fraassen. This section introduces his account and then indicates why it is safe to take it on board.

On van Fraassen's analysis, questions of the form "Why P?" are best understood as

denoting a function to sets of admissible answers, from 4-place tuples which consist of a *topic* proposition that's expressed by the interrogative, and a *contrast class* of propositions, a *relevance relation*, and a *background body* of theory and factual information that are all implicated by the conversational context in which the interrogative is uttered. We can write this as $\Theta(P_T; P_a, \ldots, P_j; R; K)$, where Θ represents the why-function, P_T represents the appropriate topic proposition presupposed by the interrogative, all $P_a, \ldots, P_j$ represent the elements of the contrast class presupposed by the interrogative, R represents relevance relation presupposed by the interrogative, and K represents the background body of theory and factual information. More colloquially, van Fraassen argues that interrogatives of the form "Why P_k?" should be understood to be expressing questions of the form "Why is it the case that P_T, with respect to R, rather than any $P_a, \ldots, P_j$, given K?" An interrogative is *appropriate* just if it expresses a why-question that enjoys admissible answers. Call any tuples that meets these constraints an *answers-apt* tuple.

The particular why-questions that van Fraassen has in mind are question types rather than question tokens. We can distinguish between token utterances of why-questions, and the propositions they express, by referring to the token utterances as interrogatives and reserving the term "question" for the propositions expressed by the interrogatives. This is an important distinction to note because, on van Fraassen's theory, the particular question that is expressed by any given interrogative will depend on the conversational context in which the interrogative is uttered.

This is admittedly quite confusing, and so I will go through these various components one at a time.

The *topic* of a why-question is the proposition that the question is about. For example, if

we ask "Why did Adam eat the apple",[31] <Adam ate the apple>[32] is the topic of the question expressed by the interrogative. Van Fraassen argues that a why-question has admissible answers only if its topic is *appropriate*. A topic is appropriate only if it is *true*. It makes no sense, on van Fraassen's view, for there to be an appropriate answer to the question why, say, Romney won the 2012 election, if <Romney won the 2012 election> is false. Van Fraassen suggests that "Why P?" questions with a false propositional content have no admissible answers, and that the suitable response will be, not an *answer*, but rather an expression to the effect that the question "Why P?" is inappropriate (a puzzled stare, a friendly correction, the statement, "your question doesn't make any sense", and so forth).

Van Fraassen introduces *contrast classes* alongside topics as a further element of the fully specified propositional objects of why-questions. They serve in part to individuate why-questions that can be expressed by syntactically identical utterances. Consider again: "Why did Adam eat the apple?" The questioner might be asking a lot of things. She might be asking why *Adam*, rather than Cain or Abel, ate the apple. She might be asking why Adam ate the *apple* rather than eating the mango or the pomegranate instead. Or she might be asking why Adam *ate* the apple rather than, say, made it into applesauce or threw it at the snake. This is an important observation, van Fraassen thinks, because each of these distinct {topic, contrast class} pairs will generate distinct sets of admissible answers. "Adam was hungry" *is* an admissible answer to the question why Adam *ate* the apple (rather than throw it at the snake). Throwing it at the snake may be very satisfying to Adam, but it will not satisfy Adam's hunger. However, "Adam was hungry" *isn't* an admissible answer to the

[31] This is van Fraassen's choice of example.

[32] I'll indicate propositions with "< >" characters.

question why Adam ate the *apple* (rather than the mango or the pomegranate). After all, any of these actions would have satisfied Adam's hunger, so Adam's hunger does nothing to answer why Adam ate the apple rather than these other things. To answer that, you'd need something like "Adam likes apples the most".

Van Fraassen generalizes this result to all why-questions. He argues that every why-question specifies a contrast class of propositions containing propositions that denote *alternative* events or states that do not occur or obtain. Further, each "Why P?" question seeks an answer as to why the topic proposition is true and why the alternatives specified by the elements of the contrast class are false. Van Fraassen does not give a rigorous analysis of alternatives. But he does state that propositions that are genuine alternatives to the topic proposition must be false (Fraassen 1980, p. 145). This is intuitive: it makes no sense to ask why Adam ate the apple rather than the mango if Adam also ate the mango.

In van Fraassen's Garden of Eden example, these alternatives are alternative actions that might have occurred instead of Adam's action of eating the apple. However, it's important to note that van Fraassen is offering the notion of contrast classes in service to a general theory of the structure of the propositional objects of why-questions, and so the contrast class in other cases might be sets of propositions expressing alternative mere events or alternative states of the world. For instance, the contrast class expressed by an interrogative like "Why is the sky blue at noon", given the context, might be the set consisting of <The sky is red at noon>, <The sky is yellow at noon>, and so on.

The relevance relation is a relation between pairs of topic questions and contrast classes, and admissible answers. It is with respect to the relevance relation that answers are deemed admissible. Depending on the conversational context of an utterance like "Why did Adam

eat the apple (rather than the mango or the pomegranate)?", the relevance relation might constrain admissible answers to ones that cite the circumstances that gave rise to Adam's eating the apple (e.g., Eve gave it to him), Adam's motives in eating it (he was hungry, he wanted to please Eve), biological factors like Adam's neural activity or the movements of his muscles, or cosmological factors like fate or providence. Consider these different answers to the question "Why did Adam eat the apple rather than the mango?"

1. Adam likes apples better than mangoes.
2. The rules of etiquette indicate that, when in Eden, one must eat apples before mangoes.
3. There was an apple rather than a mango between Adam's teeth when Adam's motor cortex produced an action potential that contracted his temporalis muscle to occlude his mandible and maxilla.
4. God willed the Fall so that humans could experience the goodness of the Incarnation, and Adam's eating the mango would not have caused the Fall.

Each of these answers will be appropriate in certain contexts and not appropriate in others. For example, if I'm in a hospital trying to save Adam's life, (3) seems perfectly appropriate. But if I answer with (3) in a theology classroom, my professor will think I'm being snarky.[33] My theology professor wants answer (4). But (4) is completely inappropriate in a hospital setting.

Here is van Fraassen's own example:

[33] In this judgment, he would also be correct.

> Suppose a father asks his teenage son, "Why is the porch light on?" and the son replies "The porch switch is closed and the electricity is reaching the bulb through that switch." At this point you are most likely to feel that the son is being impudent. This is because you are most likely to think that the sort of answer the father needed was something like: "Because we are expecting company." But it is easy to imagine a less likely question context: the father and son are re-wiring the house and the father, unexpectedly seeing the porch light on, fears that he has caused a short circuit that bypasses the porch light switch. In the second case, he is not interested in the human expectations or desires that led to the depressing of the switch. (ibid., p. 131)

Plausibly, the conversational context of the question utterance in this example specifies (<The porch light is on>, <The porch light is off>) as the contrast class. As far as the topic and contrast class is concerned, the son is giving a suitable answer to the question why the lights are on rather than off. Yet, intuitively, we want to say that, given the conversational context, the son is not giving a suitable answer to his father's question. He's merely being impertinent. Van Fraassen suggests that this is because the conversational context tells us that any admissible answer to the interrogative has to cite some feature of our human expectations and desires. Other conversational contexts could specify a feature with respect to which answers are admissible only if they cite the house's electrical properties, or various legal features (a local ordinance declaring, say, that all porches must have their lights on after dusk), perhaps the dispositions of individual electrons, and so forth.

Lastly, the background body of theory and factual information K helps to determine which of the admissible answers to the question are correct. "Adam likes apples the best" may be an admissible answer to the question "Why did Adam eat the apple rather than the mango, with respect to any features of the action involving human ends and desires", but it is not a correct answer to the question if, given K, Adam actually likes mangoes the best.

Strikingly, van Fraassen refers to correct answers to why-questions as reasons (ibid.,

p. 143) that favor (ibid., e.g. 128) the particular topic proposition (in contrast to the contrast class, with respect to the relevance relation, and given the background body of theory and factual information).

Why van Fraassen's analysis?

Van Fraassen's account of why-questions and their role in scientific explanation has been criticized on a number of fronts. Kitcher and Salmon challenge the pragmatic implications van Fraassen draws from his account. As an illustration of the pragmatic nature of scientific explanation, van Fraassen gives the famous example of the tower and the shadow. In the lawn of an aristocrat's house, a traveler sees, at sunset, a terrace cast in shadow. The traveler then notices that the shadow is being cast by a very high tower; the tower's height and location explain the presence of the shadow on the terrace. But when asking a servant why the tower had to be so high, the servant replies that it is because the aristocrat, for his own personal reasons, intended the shadow to be cast over the terrace at sunset, and constructed the tower to bring this shadow about. Van Fraassen takes both explanations to be perfectly legitimate, and offers this story to illustrate his view that whether a particular purported explanation qualifies as a genuine explanation depends on the conversational context in which the explanation is given–hence, pragmatism about scientific explanation. Kitcher and Salmon reply that it is false that the length of the shadow explains the height of the tower. Rather, what explains the height of the tower is the aristocrat's desire for the shadow, which is of course quite a different thing than the shadow itself. Kitcher and Salmon therefore argue that van Fraassen's theory of the logical structure of why-questions and their role in explanation fail to offer a genuinely pragmatic account of scientific explanation (Kitcher

and Salmon 1987), (Achinstein 1984, p. 131).

Other critics charge van Fraasson with trivializing the notion of explanation (Kitcher and Salmon 1987), of not accounting for explanation in the case of mathematics (Sandborg 1998), and as not offering a complete account of explanation because of its failure to account for the existence of some explanations that answer how-questions (Cross 1991).

These objections, along with the objections of Kitcher and Salmon above, share a theme: they are all objections to the claim that van Fraassen's analysis of why-questions is an analysis of *explanation* in general or scientific explanation in particular. Importantly, they are not objections to van Fraassen's theory of the logical structure of why-questions and their answers. Furthermore, as far as I know, there are no extant objections in the literature to van Fraassen's theory as a theory of why-questions itself.[34] Because of the lack of objections to van Fraassen's theory of the logical structure of why-questions, I will just adopt the theory and incorporate it into the argumentative structure of this dissertation.

I discussed in §2.3 Scanlon's view that the favoring relation was a 4-place relation. Specifically, Scanlon accepts

Reasons Relationality: Any reason fact is normative in virtue of its standing in

[34]Two caveats to make about this claim. First, there are rival theories of why-questions—for instance, the theory of (Hintikka and Halonen 1995) and the work on erotetic logic that was done before van Fraassen published *The Scientific Image*. What I mean is not that van Fraassen's theory of why-questions has no competitors, but that I can't find explicit objections to his theory of the logical structure of why-questions. Secondly, there does exist one small dispute about whether contrast classes ought to be included in the theory. (Ruben 1987) argues that contrast classes can be eliminated by conjoining the negations of the elements of the contrast class to the topic proposition. In other words, Ruben suggests that "Why did Adam eat the apple (rather than the mango)" is equivalent to "Why did Adam eat the apple *and not* eat the mango?" Since these are equivalent, it's better to just leave out the contrast classes. (Risjord 2000, pp. 66-78) argues that Ruben's conclusion produces inconsistencies. However, nothing in what I say in this section will hang on this issue. I argue in §5.4.3 that reasons-apt tuples need to include some proposition P such that any fact that is a reason for the agent to perform an action will be a reason for the agent to perform that action *rather than* P. So long as there's at least one such proposition, it is immaterial to my argument whether P is included within a distinct contrast class, or as a negation conjoined to the fact that serves as the reason. I will assume with van Fraassen that question-apt tuples must include a distinct contrast class.

the favoring relation with a circumstance-agent-action triple.

What that means is that the facts that are reasons are always reasons *for*: reasons for an agent to perform an action in a circumstance.

In the next two sections, I will argue that Scanlon's commitment to Reasons Relationality generates inconsistencies and that it should be rejected in favor of a more complicated view about the structure of the tuples that are apt to exemplify the favoring relation. Then, in §5.3.4, I'll argue that this more complicated view gives us reason to think that the favoring relation just is a specific instance of the relation that relates questions and answers.

Relevance relations

In this section I argue that the favoring relation includes more structure than Scanlon imputes to it. Specifically, I argue that the favoring relation needs to include a relevance relation of the sort that van Fraassen argued we needed in the case of questions and answers. In defense of this claim, I will show that Scanlon's theory of the structure of the favoring relation produces inconsistencies which can be avoided if we expand the structure of the favoring relation to include a relevance relation.

Inconsistency worries about Reasons Relationality

At the beginning of "Actions, Reasons and Causes", Davidson writes, "I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home. Here I do not four things, but only one, of which four descriptions have been given" (Davidson 1963, pp. 686-7). Davidson here is providing a component

of a theory of *action individuation*. On his view, correct descriptions are not individuation conditions of actions; or, more simply, Davidson's view is that there can be lots of correct ways to describe a single action.

Davidson gives this claim about action individuation in the service of motivating a particular view on the nature of explanatory reasons and of intentional and unintentional action. However, it also raises questions about Scanlon's theory of normative reasons. Take Sarah, whose exam is quickly approaching and who believes (correctly) that making a short run for coffee will help her prepare for it. Her upcoming exam thus intuitively provides her with a reason to run *to the nearest coffee shop* and acquire a cup of coffee. However, the nearest coffee shop also happens to be the nearest store, and it also happens to support abusive labor practices. Intuitively, Sarah's upcoming exam does not strike us as a reason for her to run *to the nearest store that supports abusive labor practices*.

Sarah's just doing one thing: going to the coffee shop, which in this case happens to support abusive labor practices. "Running to the nearest coffee shop" and "running to the nearest store that support abusive labor practices" are the same thing, the same action. Our intuitions about the reasons that Sarah has, then, generate a conflict. Intuitively, Sarah's upcoming exam gives her a reason to run to the nearest store (under the description "running to the coffee shop"). But intuitively, Sarah's upcoming exam does not give her a reason to run to the nearest store (under the description "running to the nearest store that supports abusive labor practices"). Furthermore, in both cases, we're dealing with the same agent—Sarah—and the same circumstances: whatever circumstances Sarah happens to be in. Thus Sarah's upcoming exam gives Sarah a reason in her circumstances to run to the nearest store, and Sarah's upcoming exam does not give Sarah a reason in her circumstances to run

to the nearest store. This is a contradiction.

Anscombe on action descriptions

What's going on here? I want to get a clearer view of what's going on here by considering a few of Anscombe's insights into this issue of action descriptions. Davidson, after all, got his theory of action individuation from Anscombe. Anscombe writes:

> Since a single action can have many different descriptions . . . it is important to notice that a man may know that he is doing a thing under one description and not under another. Not every case of this is a case of his knowing that he is doing one part of what he is doing and not another (e.g., he knows that he is sawing but not that he is making a squeaky noise with his saw). He may know that he is sawing a plank, but not sawing an oak plank or Smith's plank; but sawing an oak plank or Smith's plank is not something else that he is doing besides just sawing the plank that he is sawing. For this reason, the statement that a man knows he is doing X does not imply the statement that, concerning anything which is also his doing X, he knows that he is doing that thing. (Anscombe 1957, 11-12, §6)

Julia Annas suggests that, in this passage, Anscombe is expressing the view that the same action can be intentional under one description but not under another (Annas 1976, p. 252). This interpretation does not come immediately from this passage itself, but it becomes more plausible when considered in light of other things that Anscombe says. For instance, Anscombe writes, "The class of things known without observation is of general interest to our enquiry because the class of intentional actions is a sub-class of it" (Anscombe 1957, 14, §8) If, as Anscombe suggests, Jones's sawing the oak plank is the same thing as his sawing the plank, that sawing the oak plank is not something that Jones knows that he is doing (because, say, he does not know that the wood that constitutes the plank is oak rather than ash), and, third, that Jones (plausibly) is intentionally sawing the plank under the de-

scription "sawing the plank", there will be another description of the action, "sawing the oak plank", under which Jones' action is not intentional.

One might think that Leibniz's Law worries surface here. Per hypothesi, Jones' sawing of the plank and Jones' sawing of the oak plank are the same action. So Leibniz's Law should indicate that they share the same properties. But one might think that, on Anscombe's view, they don't: sawing the plank has the property of being intentionally done by Jones, while sawing the oak plank lacks this property.

Anscombe, however, explicitly argues that her view is not subject to this specific criticism:

> It is supposed by some that "x under a description d" is the form of a subject-phrase. This of course raises the question what sorts of object or entity (distinct from an A?) and A-under-the-description-d may be. But "Under a description 'putting the book down on a puddle of ink"' has as its subject simply "my action" and as predicates "intentional under a description" is "qua", or Aristotle's "hêi"[35] in modern dress. Aristotle too observes . . . that the phrase "hêi" belongs to the predicate, not to the subject. There aren't such objects as an "*A* qua *B*", though an *A* may, qua *B*, receive such and such a salary and, qua *C*, such and such a salary. . . . This is no more a rejection of Leibniz's Law than it is to say that Socrates is taller than Thaetetus and not taller than Plato". (Anscombe 1979, p. 220)

Anscombe here requests that we understand her claim that some particular intentional action is intentional under a description as a claim that the action is intentional "qua" the description. It is not entirely clear what Anscombe means by this, but her reference to Aristotle's use of the dative feminine singular form of the ancient Greek relative pronoun suggests a hermeneutic key into the passage. Aristotle begins the first part of his Metaphysics Γ by calling our attention to a type of inquiry into *tò òn hêi òn*, typically translated

[35] Anscombe uses the Greek spelling rather than the Latin transliteration that appears here.

as "being qua being"[36] or "being as being."[37] Call this "new" kind of inquiry *metaphysics**. Whereas mathematics and the other types of inquiry limit their investigations to the attributes of some particular aspect of being, the study of being qua being, for Aristotle, is one "which investigates being as being and the attributes which belong to this in virtue of its own nature".

Although Aristotle in this passage claims that he is introducing a new type of inquiry, (Cohen 2012, p. 205) argues that it is a mistake to think that he distinguishes it from the others in virtue of its investigation of a unique domain that Aristotle labels "being as being". Indeed, the subject matter of metaphysics* is, for Aristotle, the same as that of the other sciences—being simpliciter—so this shared subject matter can't serve as the individuating principle. The individuating principle that Aristotle offers for the various sciences is not these sciences' subject matter, but rather the *respect in which* the sciences study the subject matter. Metaphysics* studies being qua being, that is, it studies the attributes that belong to beings as such, insofar as they are beings. Physics, on the other hand, studies beings with respect to their motion, insofar as they are movable; mathematics studies beings with respect to their quantifiability.

All this ultimately gestures towards a plausible interpretation of the above passage from Anscombe. By claiming that actions are intentional "under such-and-such description" is to be understood as meaning that they are intentional "qua" that description, Anscombe seems to be saying that actions are intentional *with respect to* that description.

[36]E.g., the Treddenick translation

[37]e.g., Ross' translation

Reasons and relevance relations

This construal of Anscombe gestures toward an explanation of why it is that considering Sarah's action under different descriptions generates differing intuitions about whether her upcoming exam provides her with a reason to run out to the nearest shop. What is likely going on is that our consideration of the action under these distinct descriptions "makes salient" distinct features of the action and the larger circumstances against which we evaluate whether Sarah's upcoming exam provides a reason to run to the nearest shop. When we think of Sarah's action of running to the nearest shop under the description "running to the nearest coffee shop", this description makes salient the fact that there's coffee in the shop, and it's with respect to the presence of this coffee at the shop that Sarah's impending exam is a reason to run to the shop. In contrast, Sarah's impending exam does not strike us as a reason for the action when we consider the action under the labor abuse description, since this description makes salient different features of the action that have nothing to do with Sarah's exam.

This sounds strikingly similar to van Fraassen's description of relevance relations. In van Fraassen's example of the father asking the son why the lights were turned on, the presuppositional context of the father's question to his son about why the light is on makes salient certain features of the light's being on—i.e., housecalling etiquette—against which an answer that cites the imminent arrival of guests strikes us as *appropriate*. In contrast, an answer that cites the immanent arrival of the guests would not strike us as appropriate if, say, the context made clear that the father was offering his son a physics lesson right before the arrival of the guests. In Sarah's case, the presuppositional context invoked by of

our thinking of the action as running to a coffee shop makes salient features of the action against which Sarah's upcoming exam strikes us as *providing a reason* to run to the nearest shop.

Action individuation is not a worry

One might object that the conclusions I draw from the Sarah getting her coffee example are illegitimate because I am illicitly claiming that, in cases like Sarah's, the person does only one action. Perhaps, after all, events are more fine-grained than I am assuming. On a prominent view of event individuation defended by Jaegwon Kim (Kim 1969), (Kim 1976), events are uniquely referenced by event descriptions—in other words, no event can have multiple, logically non-equivalent, descriptions.[38] If it were true that Sarah is doing two actions, then there is no inconsistency problem with Scanlon's commitment to the four-place nature of the reasons relation. After all, there's no problem created by certain facts being reasons for some actions and not reasons for others. More specifically, there's no inconsistency problem if it turns out that Sarah's impending exam is a reason for the action of going to the nearest coffee shop and *not* a reason for the *distinct* action of going to the nearest shop that uses abusive labor practices. The inconsistency worry arises only if Sarah's going to the nearest coffee shop and Sarah's going to the nearest shop that uses abusive labor practices are the same thing.

I respond simply by asserting the Anscombe-Davidson thesis on action individuation and rejecting the fine-grained approach of Kim and Goldman.

There are many reasons to reject Kim's view of action individuation. Suppose first that

[38]This kind of view was also defended By Alvin Goldman in, for example, (Goldman 1971).

Bertrand assaulted Candace quickly and sharply on the kneecap (Bennett 1988, p. 78). If Kim is right, we have then a number of distinct events: Bertrand's quick and sharp assault of Candace on the kneecap, Bertrand's quick assault of Candace on the kneecap, his sharp assault of Candace on the kneecap, his sharp assault of Candace, his assault of Candace, and so forth. If events are individuated by their descriptions such that any logically non-equivalent description denoting an event does so uniquely, then Bertrand is a really, really bad guy. He assaulted Candace at least five times. It seems that a great injustice is done when the judge finds Bertrand guilty of only one count of assault.

Further, consider the tender kiss that David gave Eve on the cheek at 10:27p.m. yesterday. If events are individuated by their descriptions, this kiss is distinct from the tender kiss that David gave Eve at 10:27 yesterday. But this seems problematic. Eve, for one, would be quite upset at the suggestion. We can imagine Eve today asking David, "Why don't you ever kiss me?", and David's reply: "My dear, I've kissed you thousands of times!"[39] Secondly, though, it seems to be a metaphysical necessity that, in order to kiss someone, you have to kiss a part of them. (We say that Marilyn Monroe blows kisses at her fans, but not that she actually kisses them, except perhaps in certain special cases.) Where did David tenderly kiss Eve on the cheek at 10:27 yesterday? On the cheek. But when we ask where he tenderly kissed Eve at 10:27 yesterday, the answer is not clear. If it is indeed on Eve's cheek, it is hard to see how then we can distinguish the two kisses. Any other property—the kiss's slowness, say—would be unavailable for this purpose, as we would need to press it into service in order to individuate the tender kiss and the tender kiss on the cheek from the slow, tender kiss and the slow, tender kiss on the cheek.

[39]Eve perhaps should not be dating such an ontologically promiscuous philosopher.

Kim of course has his own set of responses to these worries, and answers the previous worry by drawing sophisticated distinction between constituting and characterizing properties. The point of including the discussion here is simply to flag some of the many serious worries about the Kimian semantics of event (action) names and the theory of event individuation in order to provide some motivation for allowing different event names to pick out the same event. Secondly, even if one accepts the fine-grained approach to action individuation, this would be consistent with my inclusion of relevance relations into the structure of the tuples that exemplify the reasons relation.

Contrast classes

In the previous section, I argued that action descriptions can sometimes make salient a subset of features of the action it describes with respect to which we evaluate whether considerations qualify as reasons for that action (for an agent, in the circumstance). There is, however, a range of cases, ones that David Enoch calls cases of arbitrary choice (Enoch 2011, pp. 73-74), that the introduction of a relevance relation does not seem to accommodate.

Here is an example of such a case. Lorenzo is at a supermarket trying to buy food for a Superbowl party. He picks up a bag of chips and now needs a can of salsa to go with them. Suppose that Lorenzo picks this particular jar of salsa because it will go well with the tortilla chips Lorenzo already picked up. Intuitively, Lorenzo does have a reason for performing the action of picking this particular jar of salsa—after all, it will go well with the chips. But if we wonder whether he has a reason to pick this particular jar of salsa rather than

one of the other twelve dozen indiscernible jars of salsa jars, the intuitive answer is: No, he doesn't have a reason to pick this particular jar rather than one of the others. They are indistinguishable in all relevant respects. So the actions are indistinguishable in all relevant respects, and so we don't have a reason to do one rather than another. Enoch goes so far as to suggest that we know this to be true (ibid., p. 74).

This intuitive tension is puzzling. These two descriptions seem to be different ways of thinking about, or describing, the very same option available to Lorenzo—grabbing this particular jar of salsa—and so it is puzzling why a consideration like the salsa's going well with the chips would appear to be a reason for electing the option when the option is considered under the first description, but not a reason for electing the option when it is considered under the second description.

Further, the introduction of a relevance relation doesn't seem to help resolve the tension present in this case. The background description of the situation makes salient the features of the option that have to do with its contributing to bringing about an enjoyable Superbowl party. It's with respect to some of these features—that picking the salsa will procure Lorenzo something that will go well with the chips, please his guests, and so forth—that the consideration is a reason for that action under the description. (The description of the situation does not move us, for instance, to consider the normative significance of Lorenzo's choosing the jar of salsa with respect to its usefulness as a doorstop or as the subject of a still-life painting.) It's true that describing the option as Lorenzo's "picking this particular jar of salsa rather than the others" highlights the salsa's proximity to a bunch of indiscernible replicas, but, unlike what we saw in the Davidson-style cases, thinking about this doesn't change our inclination to evaluate considerations as normatively salient with

respect to something other than the option's contribution to bringing about an enjoyable Superbowl party. Something else, then, must be responsible for the intuitive tension.

One possibility is that the second, more expansive description makes salient some features that constitute what Dancy calls disablers for the reason. A disabler, for Dancy, is a fact whose existence makes it the case that a consideration is not a reason. The consideration's status as a reason, in other words, depends on the nonexistence of this fact. For example, suppose that Gesenia makes a promise to Ralf to meet him for coffee at 12pm. That she made the promise gives her a reason to show up to the coffee shop at 12pm. But Ralf suddenly has to be rushed to a hospital and must miss the meeting. Gesenia no longer has the reason to show up. This is because Ralf's being rushed to the hospital disables the reason that Gesenia has in virtue of her having made the promise. A description of the action like "showing up for coffee at 12pm when Ralf had already been rushed to the hospital" would make salient a disabling feature of the situation, and so we should not expect that the promise would present itself as a reason to show up when considering the action under this description.

However, turning to our more banal salsa case, it's surely the case that the salsa's going well with the chips is still a reason to pick it, even though that jar is located next to a bunch of indiscernible replicas. The presence of these other jars doesn't disable the reason Lorenzo has to pick this particular jar. Intuitively, if we describe the action as Lorenzo's picking this particular salsa from among many indiscernible replicas, Lorenzo does have a reason to do so. He also has a reason to pick some other particular jar of salsa from among the many replicas. He just, intuitively, doesn't have a reason to pick this or that particular salsa rather than any of the others.

This may seem like it is useless hair-splitting. After all, I am drawing a distinction between "picking A rather than B, ..., N" and "picking A from among B, ..., N." Isn't this just an instance of the worst excesses of analytic intuition-pumping and concept-chopping? I think the answer is no. For one, it is unclear how any kind of philosophical work on the nature of reasons can even get off the ground without any reliance on intuitions in some form. We can debate about whose intuitions are better—those of the philosophers or those of the folk?—but we do need to appeal to intuitions at some point. As Frank Jackson is fond of saying, if we don't start from someplace intuitive, then "otherwise we start from somewhere unintuitive, and that can hardly be a good place to start from" (Jackson 1998, p. 135). Second, appealing to intuitions is a way to decide between competing theories. It is likely, particularly in ethics, that any theory that aspires to something more than being a mere laundry list of intuitively true claims about particular cases will fail to satisfy every intuition that we have. Our ethical intuitions are too muddled to be fully specified by an interesting theory. But we can accept as a general principle that a more intuitive theory is ceterus paribus better than a less intuitive one. I, naturally, think that my theory will turn out to be quite intuitive, in addition to possessing various other theoretical advantages. But we need to wait to decide whether my theory is more intuitive than others until it's fully developed.

A second possibility is that the second description highlights additional features with respect to which other considerations disfavor the action. If this were true, we should expect that the initial considerations—the salsa's going well with the chips, and so forth—would still favor the action, even though these reasons turn out to be outweighed by other, weightier reasons. Consider the description "Rachel's going into a burning building." That she

might die presents itself as a reason for her not to go in, when considered under this description. If we expand the description to highlight the fact that someone is trapped inside—say by describing it as "Rachel's going into a burning building in which two victims are trapped inside", her possible death still presents itself as a reason for her not to enter, even though the presence of the victims outweighs this other reason and ultimately gives her sufficient reason, all things considered, to enter the burning building. In contrast, when we consider Lorenzo's option under the description "picking this particular jar of salsa rather than one of the indiscernible replicas", the salsa's going well with the chips does not present itself as a reason for the option at all. So this second explanation fails to capture what is really going on in the Lorenzo case.

Does Lorenzo have a reason to pick this jar of salsa? Yes. Does he have a reason to pick this jar of salsa rather than one of the others? No. Since the introduction of relevance relations does not, by itself, resolve the intuitive tension present in these two descriptions of the Lorenzo case, we need to look elsewhere. A natural place to look would be at the differences between the two descriptions. In this case, the difference is only the inclusion of the "rather than" clause. So let's consider that.

Rather-than clauses also featured prominently in van Fraassen's theory of why-questions. We saw that, for van Fraassen, questions why some proposition is true are always, implicitly or explicitly, questions why some proposition is true rather than (in contrast to) some class of alternative, false propositions. For example, "Adam was hungry" answers the question why Adam ate the apple rather than threw the apple at the snake. But it doesn't answer the question why Adam ate the apple rather than the mango.

Analogously, we might entertain the thought that considerations that favor an action

are always, implicitly or explicitly, considerations that favor an option rather than (in contrast to) some class of alternative actions. The intuitive tension can then be explained by hypothesizing different contrast classes of options implicated when we consider the topic option under these different descriptions. For instance, "Lorenzo's choosing this jar of salsa rather than one of the others" explicitly implicates a contrast class consisting of the options of choosing this other jar of salsa, that jar of salsa over there and so on. "Lorenzo's choosing this jar of salsa" could implicitly implicate a contrast class consisting of options like Lorenzo's choosing the jar of mayonnaise (which will definitely not go well with the chips)—or, more plausibly, a contrast class consisting of an option like his not choosing anything at all.

We don't need to actually include the rather-than locution in the action description to generate the intuitive conflict. We can get the same result in just the same way that van Fraassen gets his: with emphasis. Lorenzo has a reason to choose this *jar of salsa*. But he doesn't have a reason to choose *this* jar of salsa.

All I've done so far is to indicate that there is an intuitive tension and that the introduction of contrast classes into our theory of the structure of the favoring relation would resolve this tension. However, this would not work if there were some cases of facts that were reasons not in contrast to anything at all, but were just reasons simpliciter. For example, this would be true if the fact remained a reason no matter what set of contrast classes you tested it against. One plausible candidate is something like Kantian dignity. A person's rationality gives me a reason to respect her. It also gives me a reason to respect *her* rather than someone else (since Kantian value is not comparable). And it gives me a reason to *respect* her rather than adopt any other attitude toward her (since our duty to respect rational

beings is categorical.

However, even our reason to respect a person's Kantian dignity admits of at least one minimal contrast class: every reason to respect a person's Kantian dignity will always be a reason to respect a person's Kantian dignity rather than *not* respect that person's Kantian dignity. It's a conceptual and metaphysical truth that, if you perform some action, you make it the case that you do not fail to perform that action. It's also a central feature of practical reasoning that, when we consider whether we have reasons to do something, there's a background presupposition that if we don't have reasons to do something then we won't do it. That's the whole point of considering reasons for action. If we were going to do the action anyway, irrespective of whether we have any reasons to do so, looking for reasons is a waste of time.[40] Looking for reasons to do any particular action presupposes that the reasons serve to adjudicate between doing that action and not doing that action. We can therefore sleep well at night knowing that every reason for an action is a reason for performing that action rather than not performing that action.

An aside on action omissions

I just argued that every action has a minimal contrast class consisting of the failure to perform that action. Here is an objection to this. There are no such things as failures to perform an action. Failures to perform an action are not actions; in fact, they're not anything at all. I failed to go to the moon today. That's not an action that I did. It's not anything at all. If there were such a thing, it would be a purely negative thing, a negative fact or event, and

[40]It may not be a waste of time to look for things that other people would take to be reasons with the aim of rationalizing your behavior to them. It would just be a waste of time to look for normative reasons for your behavior if you were unwilling for these reasons to impact your decision.

there are no such things as negative facts and events.

I agree that failing to go to the moon is usually nothing at all. We are failing to go to the moon all the time, and I don't want to say that there are an indefinite number of "failings to go to the moon". However, there seem to be at least a few such things—for instance, Commander John Young's failure to go to the moon as scheduled on March 17th, 1972 due to technical problems (Deffree 2015). The important difference is that, unlike our everyday failings to go to the moon, Commander Young's failure to go to the moon on that day was an intentional decision to call off the planned launch, one that Commander Young made on the basis of reasons. Intuitively, we want to say that Commander Young did do something in his failure to go to the moon.

Similarly, my ceasing to work on my dissertation at 3am seems to qualify as an action if I chose to stop working on it as opposed to simply falling asleep in my half-cubicle. And there are many similar examples. The robber walks in to the bank and shouts "don't move or I'll shoot!" He has just given you, the bank teller, a reason *not* to move, and when you comply by ceasing all movement, you have indeed done something. All these examples are examples of bona fide actions that, unlike most actions that we perform on a daily basis, are negatively defined: defined in terms of an absence of activity.

This, however, seems metaphysically problematic. How can you have something that exists but is such that all its essential features are absences? So, in this section, I'll sketch the outlines of a metaphysics of action omissions. I'll do so by applying Varzi and Casati's theory of holes to Hornsby's metaphysics of action.

Here is another example. At 9pm, I am working on my dissertation. Six hours later, at 3am, I am still working on my dissertation but I am also fighting sleep. I suggest that, given

my efforts to resist it at 3am, the omission of sleep at 3am is a positive action in a way that the omission of sleep was not a positive action (in fact, nothing at all) at 9pm when I was wide awake. Actions are a species of events. So my omission of sleep at 3am is some sort of event, and so we should look around for some sort of sleep omission event with which to identify the sleep omission action. But it seems wrong to identify this with any of the positively defined actions that I might be engaging in—the dissertation section writing, the coffee drinking and so forth. After all, these events were all present at nine pm, but the action of omitting sleep was not.

My suggested solution to the puzzle draws on an insight that I take from Jennifer Hornsby. Hornsby argues that actions are constituted by activity stuff. Sebastian's stroll through the streets of Bologna between 2 and 3pm is an action: Sebastian did stroll. This action, furthermore, is constituted by what Hornsby calls activity: unstructured event stuff that constitutes actions in a way similar to the way that an infinitely divisible clay would constitute a clay statue. Because strolling activity constitutes Sebastian's stroll, Sebastian's stroll from 2 to 3pm ontologically depends on the presence of strolling activity between 2 and 3pm in the same way that the statue ontologically depends upon its constituent clay.

Applying this to the case of my sleep omission, we could say that my positive sleep omissions are constituted by sleep omitting activity. But this alone hardly seems to help, since the notion of sleep omitting activity is no less confusing a notion than that of an action omission itself.

I suggest that a plausible account of the metaphysics of action omission can be gained if we combine Hornsby's insights with some of the insights present in a discussion of the metaphysics of holes offered by Achille Varzi and Roberto Casati (Varzi and Casati 1994).

Like action omissions, holes seem to be characterized by absence. Varzi and Casati describe a theory of holes as immaterial bodies constituted by, but not identical to, unqualified space and which ontologically depend on their material host. Call this the immaterial constitution theory. A related but distinct view of the metaphysics of holes might claim that holes are constituted not by unqualified space of the holed region but by the matter located in that region. In the case of a stereotypical donut, this second view would suggest that the donut hole is constituted not by the unqualified space of the holed region but of the material air that exists within the holed space. Call this the material constitution theory. The material constitution theory has the advantage of not needing to postulate the existence of immaterial bodies, but it runs into problems when we consider the possibility of vacuums. It seems wrong to say that the hole ceases to exist if the donut is placed in a vacuum chamber and the air previously located in the holed region is removed.

A reasonable application of the immaterial constitution theory to the case of my action of sleep omission might yield the following view. Let a ϕ-omission be an omission of an action constituted by activity of the ϕ type. For example, a sleeping omission would be an omission of an action constituted by activity of the sleeping type. Φ-omissions are actions that are constituted by unqualified spacetime and ontologically depend on the existence of an acting host and the absence of host activity of the ϕ-type in the spatiotemporal region containing the omission. In other words, in order for me to be engaging in sleep omitting activity from 3am to 3:15am, I have to exist from 3am to 3:15am and not sleeping during this time. We can be neutral for our purposes here on the precise spatial extension of the omission, but since actions are events and events have spatial extensions, it's important to include spatial extension in our characterization of action omissions.

This characterization is too liberal since, on this revised characterization, I would be engaging in sleep omission activity at 9pm as well as at 3am. After all, from 9pm to 9:15pm I also existed and was not sleeping. To solve this problem, we need to find some way of incorporating the exertion that exists at 3am and that did not exist at 9pm. After all, it was the presence of the exertion that made the difference as to the presence of the sleeping omission at 3am but not at 9pm. So let's consider this further revision. Let a phi-omission-exertion be an exertion to refrain from activity of the phi-type. Then: phi-omissions are actions that are constituted by unqualified spacetime and ontologically depend on the existence of a phi-*omission-exerting* host that enjoys an absence of host activity of the phi-type in the spatiotemporal region containing the omission.

This seems to do the trick. It also allows for the intuitive result in the following case. If I were to try to jump over a twenty-story building, we don't want to say that I am jumping over the twenty-story building. I may be trying to jump over the building but I'm not actually jumping over the building. Similarly, suppose I am struggling to stay awake and not succeeding: my eyelids are closing despite by greatest efforts to mentally will them open. In this case, although I am a sleep-omission-exerting host, nevertheless I am not omitting sleep, I am not performing the action of sleep omission, since I am falling asleep anyway and therefore failing to omit sleep. My characterization of action omissions captures this since it requires that, in addition to the host's exerting sleep-omission effort, the host must also not actually be sleeping.

There still remains the question of what phi-omissions are supposed to be. In the case of holes, we saw that the immaterial constitution theory of holes was to be preferred over the material constitution theory because it was better able to allow for the metaphysical pos-

sibility of spatial vacuums. But we need not allow for any analogous "temporal vacuum" in the case of action omissions. "Temporal vacuum" here just means time which does not contain activity, just as spatial vacuums are spaces which do not contain matter. Although there may be a need to account for temporal vacuums—if, say, we want to allow for the metaphysical possibility of an inactive but temporally extended Parmenidean monistic substance—there is no need to posit the existence of temporal vacuums in order to provide a satisfactory account of the metaphysics of action omissions. After all, action-omissions were seen to persist only across intervals during which the acting host was engaging in phi-omission-exertion activity. This activity itself minimally provides the requisite temporal stuff that we can press into service as constituting the phi-omission. Here is the result: A host's phi-omission is an action that is constituted by the host's phi-omission-exerting activity and ontologically depends on the absence of host activity of the phi-type.

Why reasons are answers

Structural similarity

Scanlon defends

> *Reasons Relationality*: Any reason fact is normative in virtue of its standing in the favoring relation with a circumstance-agent-action triple.

In the last two sections, I argued that Scanlon's theory of the four-place structure of the tuples that are apt for exemplifying the reasons relation generates inconsistencies, and that these inconsistencies can be resolved if we expand the structure through the introduction of a relevance relation and a contrast class, as follows:

> *Extended Reasons Relationality*: Any reason fact is normative in virtue of its standing in the favoring relation with a *quintuple* consisting of a fact, an agent, an action, a circumstance, a contrast class of actions, and a feature of the action with respect to which the fact is a reason.

Sarah's impending exam would not be simply a reason for Sarah in her circumstances to go to the nearest shop. Sarah's impending exam would be a reason for Sarah to go to the nearest shop rather than not go to the shop, in her circumstances, with respect to some feature of the action—say, its relevance to the value of academic achievement.

In this section, I want to argue that Extended Reasons Relationality is structurally identical to van Fraassen's view of the structure of answers-apt tuples.

On van Fraassen's view, (1) "Adam doesn't like mangoes" is an answer to why (2) Adam ate the apple, (3) rather than eating the mango, (4) given the background factual information and empirical theory, (5) with respect to our theoretical apparatus that has to do with folk psychology. Comparing this to Extended Reasons Relationality, we notice a few differences. Extended Reasons Relationality imputes six components to the reasons-apt tuples; van Fraassen imputes five components to the answers-apt tuples. Extended Reasons Relationality features a "circumstances" component, while the closest analogue in van Fraassen's answers-apt tuples is the background set of theory and factual information. The "relevance relation" in Extended Reasons Relationality is given by some sort of value such as academic achievement, while, for van Fraassen, it is given by one or another theory (folk psychology, etiquette, Christian theology, and so on). I argue that these differences are merely superficial.

Despite these superficial differences, there are important similarities between the two structures. To see this, it's helpful to observe what happens when we apply van Fraassen's theory to the question "Why did Sarah go to the nearest shop"? Suppose she did. Frequently, we explain our and other people's actions by citing the normative reasons that they have. We could also of course cite features of Sarah's psychology—we could say, for instance, that Sarah went to the nearest shop because she desired to do well on her impending exam

and believed that going to the nearest shop would help her attain that goal (because the shop has coffee, coffee has such-and-such effects on studying and so on). We could also cite features of her neurophysiology. But there is another kind of normative explanation of her action that purports to explain our actions by citing features of normative reality. Why did Sarah go to the nearest shop? Because she has reasons to do well on her upcoming exam (and the shop has coffee, and coffee has such-and-such effects on studying and so on). She went to the shop for those reasons.

Let's take this particularly normative kind of explanation as the theoretical apparatus with respect in which "Sarah has an upcoming exam" answers the question why Sarah went to the nearest shop rather than didn't go to the nearest shop, given the background set of empirical theory and factual information. Here, we have the five components that are present in van Fraassen's theory. We have a topic proposition "Sarah went to the nearest store"; a contrast class of not going to the store; a relevance relation that cites features of the particularly normative kind of explanation that I hinted at above; a set of background theory and factual information; and the answer, which in this case is "Sarah has an upcoming exam". Call the topic proposition an "event proposition", since it is a proposition about the occurrence of an event: Sarah's going to the nearest shop.

Now, let's apply Extended Reasons Relationality to Sarah's reason to go to the nearest store. Academic achievement then would be the respect in which Sarah's upcoming exam is a reason for Sarah to go to the nearest shop rather than not go to the nearest shop, given Sarah's circumstances.

The first thing to note is the similarity between the agent and action here with the topic proposition in the above statement. In the topic proposition above, we also have Sarah and we also have the action of going to the nearest shop. The agent and action tuples in Extended Reasons Relationality can be combined to form an event proposition. Thus Extended Reasons Relationality will always give us a topic proposition that meets two constraints: that the subject of the event proposition is an *agent*, and that the object of the event proposition

is an *action*. Topic propositions in a van Fraassen reasons structure don't have to consist in an agent and action, but any agent and action can be combined to form a topic proposition.

The second thing to note is that, for Scanlon, the circumstances that constitute the reasons-apt tuples were stipulated to be non-normative circumstances. I have no dispute with this feature of Scanlon's theory and agree with him that the circumstances relevant to evaluating whether a fact is a reason for an action are non-normative circumstances. Van Fraassen's background set of factual information and empirical theory is also non-normative—after all, it's *factual* information and *empirical* theory that we're talking about. Thus the circumstances component that appears in Reasons Relationality and, by extension, Extended Reasons Relationality, can itself easily constitute a background set of factual information and empirical theory.

The most significant difference between my account of reasons-apt tuples and van Fraassen's account of answers-apt tuples is our distinct accounts of the origin of the relevance relation. In van Fraassen's theory, the relevance relation is given by a specific theory: folk psychology, neuroscience, etiquette, Christian theology and so on. In my construal of the reasons-apt tuples, the relevance relation is given by a particular value with respect to which the fact is a reason for the action.

However, formally, the question of what "gives" the relevance relation is immaterial. Formally speaking, the relevance relation in van Fraassen's theory is simply a function from topic propositions and contrast classes to admissible answers. The source of that function is not itself built into the answers-apt tuple. Since Extended Reasons Relationality also includes such a function, the differences in what provides the relevance relation are not material to the question of evaluating similarities and differences between answers-apt tuples and reasons-apt tuples.

Extended Reasons Relationality's construal of reasons-apt tuples is therefore structurally identical to van Fraassen's construal of answers-apt tuples. Both give us a topic proposition, a contrast class, a relevance relation, and a circumstance.

Reasons as answers

I have just given an argument that reasons-apt tuples and answers-apt tuples have identical structures. I think, though that we should take the further step of saying not only that they are structurally identical, but that they are identical simpliciter. In particular, I suggest that reasons just are answers of a certain kind; and that the favoring relation just is a particular kind of the relation that holds between questions and answers.

In *The Scientific Image*, van Fraassen is doing philosophy of science; he is looking, in the book, for an account of scientific explanation. Thus, in developing his account of the structure of answers-apt tuples, van Fraassen considers only questions about why events occurred. Van Fraassen, however, intends his theory of questions and answers to be a theory of questions and answers simpliciter, not simply a theory of questions and answers about why certain events occur. There are other kinds of questions and answers than questions and answers about why certain events occur. Here is an example. Why *should* Sarah go to the nearest shop? Because she has an impending exam. This is not a question and answer about why a particular event occurred. It's a question and answer about why a particular event (Sarah's going to the nearest shop) *should* occur. Since van Fraassen intends his theory of questions and answers to be a theory of questions and answers simpliciter, this theory should also account for these normative-seeking sorts of questions and answers as well.

Here are two different approaches that we could take towards these normative-seeking questions and answers. On one approach, answers why an action should occur and reasons for an action to occur are distinct kinds of things. On another approach, answers why an action should occur and reasons for an action to occur are identical kinds of things. I suggest that these kinds of things are identical. Here are two arguments for this.

First: inference to the best explanation. I argued above that reasons-apt tuples and answers-apt tuples are structurally identical. Furthermore, it's always the case that, whenever a fact is a reason for an agent to perform an action, that fact will also partly answer the

question why that agent should perform that action. Why should Sarah go to the nearest shop? Because she has an impending exam. What explains this remarkable coincidence? One explanation is that, well, to be a reason for an agent to perform an action just is to be an answer the question why that agent should perform that action.

Second: parsimony. I said in §2.4 that the reasons fundamentalist held that the reasons relation was a sui generis relation that was both numerically and qualitatively distinct from other relations that we can describe in non-normative vocabulary. Lack of parsimony is typically seen as a defect of a theory, and in the case of the reasons fundamentalist, it generated serious other problems as well. We can avoid these worries about parsimony by affirming that the reasons relation is just a specific kind of the relation that obtains between questions and answers.

We need a further modification here. Sarah's impending exam does answer the question why Sarah should go to the nearest shop, but only partially. There are lots of other facts that we'd need to state in order to fully answer this question: the fact that the shop has coffee, the facts about coffee's effects on studying habits, and so on. These facts would likely be included in the circumstances against which Sarah's impending exam is a reason for her to go to the nearest shop. However, I will simply register the incompleteness of Sarah's impending exam as an answer to the question why Sarah should go to the nearest shop by stating that reasons are *partial* answers to questions why agents should perform actions:

> *Proto-Answers Account of Reasons*: Reasons are *partial correct answers* to questions why an agent should perform an action rather than not, in a particular circumstance, with respect to a relevance relation.

In §5.2, I made a great deal out of thick properties and argued that the incorporation of thick properties into the theory of reasons by means of Thick Essence of Reasons would help the irreduciblist avoid worries about Brute Entailment. The Proto-Answers Account of Reasons tells us something about what reasons are—partial correct answers to "should"

kinds of why-questions—but it does not yet say anything about thick properties. I'll discuss the role of thick properties in the next section.

5.5 Answers and thick properties

Proto-Answers Account of Reasons is a purely formal theory of reasons. It tells us something about what reasons are, and it gives us some constraints that need to be met in order for something to qualify as a reason. It is not, however, a substantive theory: it does not give us any tools to determine which facts answer which questions about what agents should do. In §5.2, I suggested that the most promising approach to this substantive issue, for the irreduciblist, is to incorporate thick properties into her theory of reasons and state that reasons are facts that bear some relation to a thick property.

Let's return again to the Sarah example. We saw that Sarah's impending exam seems to be a reason for Sarah to run to the nearest coffee shop, but seem not to be a reason to run to the nearest store that promotes abusive labor practices. These actions, per hypothesi, were the same. I also argued, through an analysis of Anscombe, that the difference between the two intuitions had to do with these distinct action descriptions' making salient distinct features of the action with respect to which the fact was a reason for the one action and not a reason for the other. To better understand the difference, then, we should better understand these distinct features of the action that the distinct descriptions make salient.

One likely possibility is that the distinct features of the action are distinct aspects of the causal profile of the action. Sarah's going to the store produces a lot of effects: it secures her some coffee, it boosts her studying performance and knowledge of the course material that she's studying, and it diminishes her bank account. It burns some calories, changes the arrangement of air molecules, and increases revenues for people that abuse their workers. Sarah's impending exam is a reason to run to the shop with respect to the action's causal properties of securing coffee and boosting her studying performance. Her

exam is not a reason to run to the store with respect to the other causal effects of the action: diminishing her bank account, crushing some blades of grass, changing the arrangement of air molecules, and increasing revenues for immoral employers.

This suggests that, at least frequently,[41] the features of an action with respect to which a fact is a reason for that action are features of that action's causal profile.

What does securing coffee have in common with boosting studying performance that it does not have in common with diminishing a bank account, crushing blades of grass, rearranging air molecules, and benefiting abusive shopowners? Securing coffee and boosting studying performance are both features of the causal profile of Sarah's running to the nearest shop. In this case, they do share a similar property: the property of directly contributing to the promotion of Sarah's performance on her impending exam. The other features of the action's causal profile also contribute in some sense to the promotion of Sarah's performance: after all, they are part of the action's causal profile. But they don't contribute in a direct way. In other words, if we imagine our wizard suspending the laws of nature and contemporary finance such that the air molecules are not rearranged, the blades of grass are not crushed, and Sarah's bank account is not diminished, the action of running to the nearest shop would still contribute to the promotion of Sarah's performance on the impending exam. Although these events do contribute to the promotion of Sarah's performance on the impending exam, they do so "accidentally" and not directly.

This suggests that the features of an action with respect to which a fact is a reason for that action are the states of affairs that that action will promote.

Similarly, the diminishing of Sarah's bank account and the benefiting of abusive shopowners have something in common: they both directly contribute to the promotion of economic injustice in a way that the other aspects of the causal profile of the action do

[41] It may not always be the case. For instance, if one is a Kantian, one might think that the feature of my action of telling the truth to Johnnie with respect to which "Johnnie is a rational agent" is a reason to tell the truth to Johnnie is the action's universalizability, which is a formal rather than a causal feature of the action. I acknowledge this position and that it may cause difficulties for the view I give here, but will not discuss it further because the dissertation is already too long.

not. If our wizard suspended the laws of nature such that the blades of grass did not bend, the air molecules did not rearrange, that Sarah's studying performance was not improved, and so on, Sarah's running to the shop would still contribute to the promotion of economic injustice. Although these other events do contribute to the promotion of economic injustice, they do so "accidentally" and not directly.

Sarah's impending exam partly answers the question why Sarah should run to the nearest shop. It partly answers this question because it tells us something about how running to the nearest shop would promote Sarah's academic performance. Interestingly, the shopkeepers' commitment to abusive labor practices tells us something about how running to the nearest shop would promote economic injustice. However, the shopkeepers' commitment to abusive labor practices does not answer the question why Sarah should run to the nearest shop. It doesn't even answer the question why Sarah should run to the nearest shop with respect to the action's consequences for economic injustice. What is the difference here? Plausibly, the difference has to do with the fact that Sarah's academic performance is *good*, and economic injustice is *not good*. The shopowners' abuse of workers answers the question why Sarah's running to the nearest shop would promote injustice. However, the shopowners' abuse of workers is not a reason for Sarah to run to the nearest shop because the shopowners' abuse of workers does not answer any question as to why Sarah's running to the nearest shop promotes a good state of affairs. It therefore does not answer any question as to why Sarah *should* run to the nearest shop.

This suggests that the features of an action with respect to which a fact is a reason for that action are *good* states of affairs that that action will promote. In other words, a fact can be a reason for an action only with respect to some good state of affairs that that action will promote.

With this, we have secured a role for thick properties and an argument for why they need to be incorporated into our theory of the nature of reasons. Take a property like academic achievement. Sarah's upcoming exam is a reason to run to the nearest store with respect

to the action's promotion of academic achievement. In other words, Sarah's running to the nearest store will bring about the state of affairs of Sarah's academic achievement, and Sarah's upcoming exam is a reason for running to the store with respect to this value because it partly answers the question why running to the store will bring about the state of affairs of Sarah's academic achievement and therefore promote the property of academic achievement. However, we saw that facts can be reasons for action only with respect to states of affairs that are *good*. Thus, Sarah's upcoming exam is not only a reason for running to the store with respect to the property of academic achievement which running to the store will promote. It's a reason for running to the store with respect to the property of "*academic achievement* $\wedge$ *good*", which running to the store would promote. But the property "academic achievement $\wedge$ good" is a thick property, since it contains both a non-normative element (the academic achievement) and a normative element (the goodness).

The result is this:

> *Answers Account of Reasons**: Reasons are *partial correct answers* to questions why an agent should perform an action rather than not, in a particular circumstance, with respect to a (good) thick property that the action would promote.

The qualifier that the thick property be "good" is there in order to capture the earlier observation that facts are never reasons with respect to the promotion of "bad" thick properties.

5.6 Thick properties and Thin Centrality

Answers Account of Reasons* is in tension with one of the components of my theory that I stated at the very beginning of this chapter:

> *Reasons Thin Centrality*: For all facts F, F is a *thin* normative fact only if F is a reasons fact or F embeds a reasons fact.

The tension is due to the inclusion of the thin property goodness in Answers Account of Reasons*. Answers Account of Reasons* analyzes reasons in terms of goodness. However,

Reasons Thin Centrality states that all thin normative facts, such as the fact that such-and-such thick property is good, to be facts about the instantiation of the reasons relation. So the fact that such-and-such thick property is good would have to be a fact about something's being a reason—that is to say, a fact about something's promoting something that is good. This is circular.

One option here is to hold that this circle is virtuous. It's not inconsistent to say that academic achievement is good because academic achievement has features that provide reasons to promote academic achievement, and that these features provide reasons to promote academic achievement in virtue of academic achievement's goodness. This would be a kind of distinctly metaethical coherentism: a coherentism with respect to the meaning of our normative concepts. It would be odd and perhaps unusual, but it's not inconsistent. Our concepts of goodness, of being a reason, and of being a thick property would form what Peacocke calls a "micro-holism", and it's not completely beyond the pale that such a micro-holism might be a feature of our normative concepts.

I do not myself favor this metaethical coherentist approach to reconciling Answers Account of Reasons* and Reasons Thin Centrality. The solution that I favor bears similarities to Augustine's theory of normativity. Augustine, as a theist, is very concerned that God not be responsible for the creation of evil and the failure to stop evil from occurring. His (Augustine's not God's) solution to this problem is to deny evil's existence. Evil, for Augustine, is not a property. It is a property *absence*—specifically, an absence (or privation)[42] of goodness. Because of this, Augustine argues that God cannot be responsible for creating evil or failing to stop it from occurring. After all, evil does not exist. It's just a lack of goodness. When we say that something is evil or an evil, Augustine thinks that we are saying, rather, that that thing is less good than it could be, or less of a good than it could be. But there's nothing wrong with God's having created a world that's not the best world

[42]For Augustine and Descartes, privations are absences of something that ought to be the case. I lack a third hand. This is an absence. But it's not a privation, since it should not the the case that I have a third hand.

that God could have created. The world is still good! What are you complaining about? Be thankful that you have a world at all!

I suggest that something like this is true in the case of "bad" thick properties like economic injustice and cruelty. I suggest that economic injustice, for example, is not a thick property. It is the privation of a thick property—in this case, the privation of full economic justice. Cruelty would be a privation of benevolence, dishonesty a privation of honesty, and so on.

If we accept that there are no such things as "non-good" thick properties, we can take out the "good" qualifier from our Answers Account of Reasons*. After all, if there are no non-good thick properties, then qualifying that the thick property that features in our account of reasons be a good thick property would be redundant. The result is

> *Answers Account of Reasons*: Reasons are *partial correct answers* to questions why an agent should perform an action rather than not, in a particular circumstance, with respect to a thick property that the action would promote.

I have just described two ways in which the buck-passing view of thin normative properties can be reconciled with my view that reasons are essentially constituted by thick normative properties. These two ways were: micro-holism about normative properties such that thick properties were also constituted by thin; and a foundationalist view that thick normative properties were the fundamental constitutent of normative reality. There are virtues to each approach.

In one sense, the thick properties should be considered the fundamental constituents of normative reality because they are the fundamental constituents of our normative lives. Considerations about thick normative property instantiations help us get around in distinct social contexts. For example, considerations about honor can help people navigate situations like that of the Wild West or the 19th century Russian military, and considerations about holiness can help people navigate social situations like those that exist in some seminaries or monasteries. Thin properties are much less able to help us navigate these social

situations because they lack descriptive content and are therefore determinations of their distribution could turn out to be more difficult.

However, this is not to say that thin properties are irrelevant to our normative lives. Thin properties become particularly relevant when thick properties conflict, as they serve the role of providing a single set of ranking criteria by abstracting from the descriptive content of the conflicting thick properties. Williams' Gauguin may have had difficulty deciding between being loyal to his wife or audaciously sailing to Tahiti, but this difficulty may have been more easily resolved had Gauguin considered the relative goodness of each action and chosen the better of the two. If thick normative properties like loyalty and audacity did have a goodness component such that this component could be isolated in each property and compared in this way, the result, when combined with the buck-passing account of goodness and the thick property account of reasons, would be a micro-holism: thick properties constituted by thin properties like goodness, thin properties constituted by reasons, and reasons constituted by thick properties.

I myself do not take a stance on which view would best reconcile my thick property account of reasons with a buck-passing view of thin properties. I do, though, take this to be a substantially underexplored question worth pursuing in significantly more detail.

5.7 A note on thick properties and normative motivation

I want to end this dissertation with some remarks on normative motivation. Normative motivation is the branch of moral psychology that studies whether and how our thoughts that some action is good or bad, right or wrong, or what we ought or ought not to do all things considered, motivate our ϕ-ing or intending to ϕ or not ϕ, or otherwise feature in our ϕ-ing or not ϕ-ing. More simply, normative motivation is the study of how our thinking that we ought to do something relates to our doing it.

Usually this branch of moral psychology is called *moral motivation* rather than norma-

tive motivation, but I agree with Rosati that normative motivation is the more accurate term to describe the study of how thoughts that something is a reason, whether moral or not, can motivate our actions (Rosati 2006).

I have not discussed normative motivation up to this point, first, because questions of normative motivation are outside the purview of normative metaphysics, the topic of my dissertation. Second, normative motivation is conceptually secondary to normative metaphysics. One could agree with reasons fundamentalists that there are irreducibly normative truths about the exemplification of the irreducible favoring relation that supervenes on the non-normative features of the world and ground all normative truths, but also consistently tell a different story about moral psychology and normative motivation than the one that Scanlon and the other reasons fundamentalists (Parfit, Enoch) tell. What follows in the next paragraphs are some remarks to clarify my claim here.

Normative metaphysics is theoretically prior to normative motivation

The fact that the favoring relation has the features that reasons fundamentalists say it does is consistent with our not being motivated to do anything for reasons at all. Suppose that all agents were maximally akratic: although they reasoned correctly and came to correct judgments about what they ought to do, these judgments never translated into action–perhaps because other mental events get in the way; perhaps because, as Frankfurt once described in a thought experiment, we are all hypnotized in such a way as to neurologically stymie the realization of any intentions to act that might form (Frankfurt 1988, p. 7).

Secondly, suppose that practical reason itself could never result in correct judgments about what one ought to do all things considered. Shakespeare offers a helpful description of what this might be like in this following monologue, spoken by the clinically depressed Hamlet:

> And thus the Native hue of Resolution
> Is sicklied o'er, with the pale cast of Thought,

And enterprises of great pitch and moment,
With this regard their Currents turn awry,
And lose the name of Action. (Shakespeare 2015)

The examples of maximally akratic agents and agents in the tragic grip of a practical reason that can never generate rational intentions to act reinforce that the question of what reasons are can be answered, if it can be answered, independently of the question whether and how these reasons motivate us. In particular, the question of normative motivation seems to me to be an empirical matter which we would be able to resolve given a theory of normative epistemology—that is, how reasons themselves factor into the formation and reliability of our beliefs about them. Furthermore, any development of the theory of normative knowledge is going to require a prior development of the theory of the nature of reasons. In saying this, I am claiming, as Peacocke does in the case of consciousness and the first person, that the metaphysics of the favoring relation enjoys explanatory and theoretical priority over the epistemology of the favoring relation (Peacocke 2014, p. 2).

Since normative motivation is conceptually posterior to normative metaphysics, we are in a better position to discuss normative motivation now that the normative metaphysics is done.

Humean and non-Humean normative motivation

Reasons fundamentalism agrees with Humean views in claiming that many of our actions are motivated by reasons. They differ with Humean views, however, in their stance on what reasons are, which ultimately generates substantial differences between the Humean and reasons fundamentalist theories of rational motivation.

On one kind of Humean model of reasons, reasons are *primary reasons*–that is, belief-desire pairs.[43] When these belief-desire pairs interact in the appropriate way, this inter-

[43] Davidson uses the term "primary reason" in a somewhat more expansive way, to include all pairs of pro-attitudes and beliefs–pro-attitudes being a more expansive category than the category of desire. (Davidson 1963, p. 686).

action, for the Humean, causes an action. Something motivates an agent to do something just if it "gets" that agent to do that thing, in some way. So reasons, for the Humean, can motivate their agents to perform actions.

For instance, Smith drinks the petrol. Why did Smith do that? A Humean might answer that Smith wanted to drink alcohol and he believed that the drink in the glass before him was a gin and tonic. Smith's belief that the liquid was a gin and tonic interacted with Smith's desire for alcohol to cause Smith's action of drinking the liquid in front of him (which in this case happens, unfortunately for Smith, to be petrol) (Williams 1981, p. 102). Since Smith's belief that the liquid was gin, and desire to drink alcohol, "got" Smith to drink the liquid in front of him, the belief-desire pair motivated Smith to drink the petrol. And since these belief-desire pairs are reasons, the Humean I am describing has a theory on which reasons can motivate us to act.

Contemporary Humeans about practical reason then to reject that beliefs constitute reasons. Contemporary Humean Mark Schroeder provides, for example, the following desire-based (subjectivist) analysis of reasons:

> For [some proposition] *R* to be a reason for [an agent] *X* to do [action] *A* is for there to be some [state of affairs] *p* such that *X* has a desire whose object is *p*, and the truth of *R* is part of what explains why *X*'s doing *A* promotes *p*. (Schroeder 2007, p. 59)

Schroeder's subjectivist analysis of reasons makes no reference to beliefs because he wants to allow for agents to have reasons even in cases where they are ignorant of what their desires are or how to bring them about. Most of us want to say, for instance, that Lonnie has reasons to drink water rather than milk when he risks being dehydrated in the desert (assuming Lonnie desires to be sufficiently hydrated) even if he lacks the belief that milk will dehydrate him (Railton 1986, pp. 174-175). Schroeder's analysis allows for the water's propensity to hydrate better than milk to be a reason for Lonnie to drink the water rather than the milk: drinking the water promotes Lonnie's desire for the state of affairs of his being hydrated to obtain. If Lonnie drinks the milk rather than the water due to his false belief

that milk better hydrated him, then Lonnie would be motivated by a belief in a proposition that is not a reason on Schroeder's analysis: it's not true that Lonnie's drinking milk would hydrate Lonnie, and therefore the proposition that Lonnie's drinking milk would hydrate Lonnie cannot feature in an explanation of why drinking milk would promote Lonnie's desire to be hydrated. However, Schroeder's analysis also allows for reasons to motivate us in the case that we do believe the reason and act accordingly. Water's propensity to hydrate Lonnie is a reason for Lonnie to drink water, given Lonnie's desire to be hydrated. Therefore, if Lonnie did believe that water hydrated him better than milk, and drank water on the basis of this consideration, Lonnie would be motivated to drink water for the reason that water would hydrate Lonnie. So even sophisticated Humean accounts of reasons such as Schroeder's account allow for reasons to motivate agents to perform actions.

Reasons fundamentalists tend to agree with the Humean that reasons motivate their agents to perform actions. But because they disagree with the Humean about what reasons are, they end up telling a different story as to what this motivation consists in. Scanlon, for instance, writes that the Humean reduction of reasons to belief-desire pairs "is refuted by the evident lack of intrinsic normative significance of facts about desires" (Scanlon 2014, p. 7). Scanlon agrees that one's desire for some state of affairs to obtain can motivate agents to attempt to bring about that state of affairs, but he argues that the mere existence of a desire for some state of affairs to obtain does not suffice to provide one with reasons to bring about that state of affairs.[44] Scanlon is also concerned that subjectivist theories of reasons cannot explain the normativity of one's reasons to conform one's actions to the subjectivist maxim that one ought to do what best promotes one's desires.

Scanlon's rejection of the Humean view of the nature of reasons leads Scanlon to hold, again contrary to the Humean, that one can adequately answer the question of rational mo-

[44]Scanlon therefore lacks Schroeder's intuition that "there is a reason for everyone to do what promotes her desires" (Schroeder 2007, p. 55). Schroeder analyzes this claim as the claim that, for all agents and actions, if an agent's doing some action promotes one or more of that agent's desires, then there is a reason for that agent to do that action. I also lack this intuition, for what it's worth, so it is probably too optimistic for Schroeder to claim that this claim is "obviously true".

tivation without referencing desires. He writes,

> a rational person who judges there to be compelling reason to do A normally forms the intention to do A, and this judgment is sufficient explanation of that intention and of the agent's acting on it (since this action is part of what such an intention involves). There is no need to invoke an additional form of motivation beyond the judgment and the reasons it recognizes, some further force to, as it were, get the limbs in motion. (Scanlon 1998, pp. 33-34)

Scanlon is not denying that desires may play a role in motivation. I think, for instance, that Scanlon would agree that desires would appear in an adequate explanation of actions done by people in the grip of extreme passion. These actions, however, would not count as actions done *for reasons*.[45] And we can give sufficient explanations of actions that agents do when acting *for reasons* simply with references to that agent's judgments about what that agent has most reason to do.

I would like to compare Scanlon's view with a distinct view of normative motivation that one might have. On one view of normative motivation, normative properties motivate us to action not only through our forming judgments in response to them, but also perceptions of them. The idea is that, when we see some kids burning a cat, we not only judge the presence of cruelty but actually see the cruelty.

Tamar Szabo-Gendler has recently argued that, in addition to beliefs, there are more primitive, belief-like states called *aliefs* that are representative and behavioral, and that govern us in cases of instinctual response. If you see a rope, perceive it as being a snake, and then jump in response, Szabo-Gendler argues that your perception and your jumping were mediated by the formation and activity of an alief that both represents the rope as a snake and motivates, in a sub-personal way, your jumping.

[45] This forms the basis of another argument that the reasons fundamentalist could make against the Humean. Intuitively, someone who commits a violent crime in the grip of passion is motivated by desires. But these desires do not provide us with reasons to commit crimes. Jones' desire to control everything about his spouse does not make his spouse's having gone out with her friends without telling Jones into a reason for Jones to hit her. One could simply deny that the motivating mental events in cases of action from extreme passion are desires. Scanlon and Quinn worry that this approach is unlikely to succeed, since the only say to distinguish the bona fide desires from the affective attitudes that aren't would involve appeal to normativity, and therefore to the notion of a reason (Scanlon 1998, pp. 38-39), (Quinn and Foot 1993, p. 228).

Aliefs share similarities with mental states that have sometimes been called "besires". A besire is a mental state that is both belief-like and desire-like. Besires are belief-like in that they represent the world as being a certain way, and they are desire-like in that they are capable of motivating. Although there is no clear consensus on what exactly a besire is in the literature, besires are usually understood to represent the world at the level of judgment. Thus besires are less primitive, more sophisticated mental states than aliefs, despite sharing similarities with them.

Scanlon's position, expressed in the blockquote above, is that the presence of an agent's judgment that X is sufficient reason to do Y provides sufficient explanation as to why that agent did Y; there is no need to further appeal to any additional form of motivation beyond the judgment. Scanlon seems to be drawn to this externalism about normative motivation from a strong conviction that desires do not provide reasons, a conviction that he develops at length in chapters 1 and 2 of his *What We Owe to Each Other*.

It is possible to share with Scanlon this conviction that desires do not provide reasons, but avoid Scanlon's commitment to the view that judgments alone can motivate. The reason is simple: the furniture of the mind relevant to motivation includes much more than beliefs, desires, intentions and actions. It might also include besires and aliefs: states which are not desires and do not decompose into desires, but yet are affective in nature, making it more easily comprehensible how such states could "get the limbs in motion". That Scanlon seems to fail to consider this possibility may be a consequence of his rejection of Humean theories of normative motivation, which understand motivation as produced by what Davidson called a "primary reason": a belief-desire pair. Scanlon rejects this theory of normative motivation by refusing to countenance that desires have a role in normative motivation. But, being immersed in the Humean model which itself recognizes only beliefs and desires as salient to motivation, Scanlon is left only with judgments to explain action. Enriching our view of mental states to include aliefs and besires may provide accordingly enriched accounts of normative motivation and of agency, neither which I am, unfortunately, able to

give here.

Bibliography

Achinstein, Peter (1984). "The Pragmatic Character of Explanation." In: *Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1984, pp. 275–292.

Anderson, Elizabeth (1990). "Is Women's Labor a Commodity?" In: *Philosophy and Public Affairs*, pp. 71–92.

Annas, Julia (1976). "Davidson and Anscombe 'On the Same Action'." In: *Mind* 85.338, pp. 51–257.

Anscombe, G.E.M. (1957). *Intention*. Harvard University Press.

— (1979). "Under a Description." In: *Nous* 13, pp. 219–233.

Armstrong, David (1978). *A Theory of Universals: Universals and Scientific Realism Vol. II*. Cambridge University Press.

— (1989). *Universals: An Opinionated Introduction*. Westview Press.

— (1997). *A World of States of Affairs*. Cambridge University Press.

Ayer, AJ (1952). *Language, Truth and Logic*. Dover.

Baldwin, Tom (1990). *G. E. Moore*. Routledge.

— (2010). "Comments on A. K. Bilgrami's Self-Knowledge and Resentment." In: *Philosophy and Phenomenological Research* 81, pp. 773–782.

Baylis, Charles A. (1948). "Facts, Propositions, Exemplification and Truth." In: *Mind* 57, pp. 459–479.

Beebee, Helen (2006). *Hume on Causation*. Routledge.

Bennett, Jonathan (1988). *Events and their Names*. Hackett.

Bilgrami, Akeel (2006). *Self-Knowledge and Resentment*. Harvard University Press.

Black, Max (1952). “The Identity of Indiscernibles.” In: *Mind* 61.242, pp. 153–164.

Blackburn, Simon (1993). “Essays in Quasi-Realism.” In: Oxford University Press. Chap. Moral Realism, pp. 111–129.

Blackburn, Simon and Allan Gibbard (1992). “Morality and Thick Concepts.” In: *Proceedings of the Aristotelian Society* 66, pp. 285–299.

Block, Ned (1978). “Troubles with Functcionalism.” In: *Minnesota Studies in the Philosophy of Science* 9, pp. 261–325.

Boyle, Matthew (2009). “Two Kinds of Self-Knowledge.” In: *Philosophy and Phenomenological Research* 78, pp. 133–164.

— (2016). “Additive Theories of Rationality: A Critique.” In: *European Journal of Philosophy* 23.

Brandom, Robert (1998). *Making it Explicit*. Oxford.

Brink, David (1984). “Moral Realism and Skeptical Arguments from Disagreement and Queerness.” In: *Australasian Journal of Philosophy*, pp. 111–125.

— (1989). *Moral Realism and the Foundations of Ethics*. Cambridge University Press.

— (2001). “Realism, Naturalism and Moral Semantics.” In: *Social Philosophy and Policy*, pp. 154–176.

Broome, John (2004). “Reason and Value: Themes from the Philosophy of Joseph Raz.” In: Clarendon Press. Chap. Reasons.

— (2013). *Rationality through Reasoning*. Blackwell Publishing.

CNN (2014). “JPMorgan’s Dimon gets 74 percent pay hike despite legal woes.” English. In: *CNN*. URL: `http://money.cnn.com/2014/01/24/news/companies/dimon-pay/index.html`.

Chalmers., David (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, David (2002a). “Conceivability and Possibility.” In: Oxford University Press. Chap. Does Conceivability entail Possibility? Pp. 145–200.

Chalmers, David (2002b). “Consciousness and its Place in Nature.” In: *consc.net*. URL: `http://consc.net/papers/nature.html`.

Chalmers, David (2004). “Soames’ Epistemic Two-Dimensionalism.” In: *Unpublished*.

Chappell, Tim (2014). “A non-natural reason by any other name...” In: *Unpublished*. URL: `http://philpapers.org/rec/CHAANR-2`.

Cohen, S Marc (2012). “The Oxford Handbook of Aristotle.” In: Oxford. Chap. Alteration and persistence: Form and matter in the physics and de generatione et, p. 205.

Copp, David (2003). “Why Naturalism?” In: *Ethical Theory and Moral Practice*, pp. 179–200.

Correia, Fabrice and Kevin Mulligan (2013). “Stanford Encyclopedia of Philosophy.” In: Stanford University. Chap. Facts.

Cross, Charles B (1991). “Explanation and the Theory of Questions.” In: *Erkentniss* 34.2, pp. 237–260.

Cullison, Andrew (2010). “Moral Perception.” In: *European Journal of Philosophy* 18:2, pp. 159–175.

Dancy, Jonathan (2004). *Unprincipled Ethics*. Oxford University Press.

Davidson, Donald (1963). “Actions, Reasons and Causes.” In: *The Journal of Philosophy* 60, pp. 685–700.

— (2001). *Essays on Action and Events*. UC Berkeley Press.

Deffree, Suzanne (2015). “Apollo 16 launches after month-long delay, April 16th 1972.” In: *EDN Network*.

Dreier, Jamie (2006). “Metaethics after Moore.” In: Oxford. Chap. Was Moore a Moorean? Pp. 191–207.

— (2015). “Motivational Internalism.” In: Oxford University Press. Chap. Can Reasons Fundamentalism Answer the Normative Question? Pp. 167–180.

Egan, Andrew (2007). “Quasi-realism and fundamental moral error.” In: *Australasian Journal of Philosophy* 85, pp. 205–219.

Engel, Nicholas (2015). “Reply to Jules Salomone, “The Ontology of Money”.” In: *15th Annual NYU-Columbia Graduate Student Conference*.

Enoch, David (2011). *Taking Morality Seriously*. Oxford.

Ewing, A. C. (1939). “A Suggested Non-Natural Analysis of “Good”.” In: *Mind* 48, pp. 1–22.

Field, Hartry (2001). “Peacocke and Boghossian 2001.” In: Oxford University Press. Chap. Apriority as an Evaluative Notion, pp. 117–149.

Fine, Kit (1994). “Essence and Modality.” In: *Philosophical Studies*.

— (2001). “The Question of Realism.” In: *Philosopher’s Imprint* 1.

— (2002). “Conceivability and Possibility.” In: Clarendon Press. Chap. Varieties of Necessity, pp. 263–291.

— (2012). “Metaphysical Grounding: Understanding the Structure of Reality.” In: Cambridge University Press. Chap. A Guide to Ground, pp. 37–81.

Foot, Phillippa (1967). “The Problem of Abortion and the Doctrine of Double Effect.” In: *Oxford Review* 5.

Fraassen, Bas van (1980). *The Scientific Image*. Oxford.

Frankena, William (1963). *Ethics*. Prentice-Hall.

Frankfurt, Harry (1988). *The Importance of What We Care About and Other Essays*. Cambridge University Press.

Geach, Peter (1956). “Good and Evil.” In: *Analysis* 17, pp. 33–42.

Gibbard, Allan (1990). *Wise Choices, Apt Feelings*. Harvard University Press.

— (2003). *Thinking How to Live*. Harvard University Press.

Goldman, Alvin I. (1971). “The Individuation of Action.” In: *The Journal of Philosophy* 68.21, pp. 761–744.

Greene, Joshua (2007). “Moral Psychology: Historical and Contemporary Readings.” In: MacMillan. Chap. The Secret Joke of Kant’s Soul, pp. 259–372.

Hare, R. M. (1982). *Moral Thinking*. Oxford.

— (1984). “Inaugural Address: Supervenience.” In: *Proceedings of the Aristotelian Society* 58, pp. 1–16.

Hieronymi, Pamela (2005). “The Right Kind of Reason.” In: *The Journal of Philosophy* 102.9, pp. 437–457.

— (2013). “The Use of Reasons in Thought.” In: *Ethics* 124, pp. 114–127.

Hintikka, Jaakko and Ilpo Halonen (1995). “Semantics and Pragmatics for Why-Questions.” In: *The Journal of Philosophy* 92.12, pp. 636–657.

Hirsch, Eli (2011). *Quantifier Variance and Realism: Essays in Metaontology*. Oxford.

Ho, Tracy (2009). *Most Important Philosophical Books since 1950*. Leiter Report. URL: `http://leiterreports.typepad.com/blog/2009/05/the-most-important-philosophical-books-since-1950.html`.

Horgan, Terence and Marc Timmons (1991). “New Wave Moral Realism Meets Moral Twin Earth.” In: *Journal of Philosophical Research* 16, pp. 447–465.

Hornsby, Jennifer (2012). “Actions and Activity.” In: *Philosophical Issues* 22.1, pp. 233–245.

Hurley, Susan (1989). *Natural Reasons*. Oxford.

Jackson, Frank (1998). *From Metaethics to Ethics*. Oxford University Press.

Jackson, Frank and Phillip Pettit (1993). “Folk Belief and Commonsense Belief.” In: *Mind and Language* 8, pp. 298–305.

Johnston, Mark (1989). “Dispositional Theories of Value.” In: *Proceedings of the Aristotelian Society* 63, pp. 139–174.

Kaspar, David (2012). *Intuitionism*. Bloomsbury.

Kearns, Stephen and Daniel Starr (2008). “Reasons: Explanations or Evidence?” In: *Ethics* 119, pp. 31–56.

Kim, Jaegwon (1969). “Essays in Honor of Carl G. Hempel.” In: Springer. Chap. Events and their Descriptions: Some Considerations, pp. 198–215.

— (1976). “Events as Property Exemplifications.” In: *Action Theory: Proceedings of the Winnipeg Conference on Human Action, Held at Winnipeg, Manitoba, Canada, 9–11 May 1975* Synthese Library 97, pp. 159–177.

— (1987). “Supervenience and Mind: Selected Philosophical Essays.” In: Cambridge University Press. Chap. “Strong” and “Global” Supervenience, pp. 79–91.

— (1989). “Mechanism, Purpose, and Explanatory Exclusion.” In: *Philosophical Perspectives* 3, pp. 77–108.

— (1998). *Mind in a Physical World*. Cambridge University Press.

— (2005). *Physicalism, or Something Close Enough*. Oxford University Press.

Kitcher, Philliip and Wesley Salmon (1987). "Van Fraassen on Explanation." In: *The Journal of Philosophy* 84.6, pp. 315–330.

Kitcher, Phillip (1983). *The Nature of Mathematical Knowledge*. Oxford University Press.

Korsgaard, Christine (1981). "The Standpoint of Practical Reason." PhD thesis. Harvard University.

— (1996). *The Sources of Normativity*. Cambridge University Press.

— (2009). "Normativity, Necessity, and the Synthetic A Priori: A Response to Derek Parfit." URL=<http://philpapers.org/rec/KORNNA>.

Kripke, Saul (1980). *Naming and Necessity*. Blackwell Publishing.

LaPorte, Joseph (2004). *Natural Kinds and Conceptual Change*. Cambridge University Press.

— (2006). "Rigid Designation." English. In: *Stanford Encyclopedia of Philosophy*. The Stanford Encyclopedia of Philosophy Summer 2011 Edition. Ed. by Joseph LaPorte. URL: http://plato.stanford.edu/archives/sum2011/entries/rigid-designators/.

Leiter, Brian (2010). *The Most Significant Moral Philosophers of the Twentieth Century: The Results*. Leiter Report. URL: http://leiterreports.typepad.com/blog/2010/10/the-most-significant-moral-philosphers-of-the-20th-century-the-results.html.

Lenman, James (2006). "Moral Naturalism." In: *Stanford Encyclopedia of Philosophy* Spring 2014 Edition. URL: http://plato.stanford.edu/archives/spr2014/entries/naturalism-moral/.

Levi, Isaac (1986). *Hard Choices*. Cambridge University Press.

Levi, Primo (1993). *Survival in Auschwitz: The Nazi Assault on Humanity*. Simon and Schuster.

Levin, Janet (2013). "Functionalism." English. In: *Stanford Encyclopedia of Philosophy*. URL: http://plato.stanford.edu/archives/fall2013/entries/functionalism/.

Lewis, David (1973). *Counterfactuals*. Oxford University Press.

— (1986). *On the Plurality of Worlds*. Blackwell Publishing.

— (1991). *Parts and Classes*. Blackwell Publishing.

— (2000). "Causation as Influence." In: *Journal of Philosophy* 97, pp. 182–197.

Macfarquhar, Larissa (2011). "How to be Good." In: *The New Yorker* September 5, 2011. URL: `http://www.newyorker.com/magazine/2011/09/05/how-to-be-good`.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. Penguin.

Majors, Brad (2009). "Oxford Studies in Metaethics." In: vol. 4. Oxford University Press. Chap. The Natural and the Normative, pp. 29–52.

McDowell, John (1998). "Non-Cognitivism and Rule-Following." In: *McDowell 1998*, pp. 198–220.

McLaughlin, Brian (1995). "In Savellos 1995." In: Cambridge University Press. Chap. Varieties of Supervenience, pp. 16–59.

McNaughton, David (1988). *Moral Vision*. Oxford: Basil Blackwell.

Melnyk, Andrew (2003). *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge University Press.

Merricks, Trenton (1998). "Against the Doctrine of Microphysical Supervenience." In: *Mind* 107, pp. 59–71.

— (2003). "Maximality and Consciousness." In: *Philosophy and Phenomenological Research* 66, pp. 150–158.

— (2007). *Truth and Ontology*. Oxford.

Meyers, Chris (2012). "Expressivism, Constructivism, and the Supervenience of Moral Properties." In: *Philosophical Explorations* 15, pp. 17–31.

Moffett, Marc (2010). "Against A Posteriori Functionalism." In: *Canadian Journal of Philosophy* 40, pp. 83–106.

Moore, GE (1922). "The Conception of Intrinsic Value." In: *Philosophical Studies*, pp. 260–275. URL: `http://ecaths1.s3.amazonaws.com/topetica/The.Conception.of.Intrinsic.Value.1670334838.pdf`.

Moore, G.E. (1993). *Principia Ethica, Revised Edition*. Cambridge University Press.

Mothersill, Mary (1952). "Moral Philosophy and Meta-ethics." In: *The Journal of Philosophy* 49, pp. 587–594.

Mourelatos, Alexander (1978). "Events." In: Dartmouth: Aldershot. Chap. Aristotle's Kinesis/Energeia distinction: a marginal note on Kathleen Gill's paper, pp. 415–434.

Nagel, Thomas (1978). *The Possibility of Altruism*. Princeton University Press.

Nietzsche, Friedrich (2012). *Daybreak*. Gutenberg. URL: `http://www.gutenberg.org/files/39955/39955-h/39955-h.html`.

Nolan, Daniel (1997). "Impossible Worlds: A Modest Approach." In: *Notre Dame Journal of Formal Logic* 38, pp. 535–572.

Normore, Calvin (2010). "Fool's Good and Other Issues: Comments on Self-Knowledge and Resentment." In: *Philosophy and Phenomenological Research* 81, pp. 766–772.

Oddie, Graham (2009). *Value, Reality, and Desire*. Oxford.

Olson, Jonas (2012). *Moral Error Theory: History, Critique, Defense*. Oxford University Press.

Parfit, Derek (1998). "Why Anything? Why This?" In: *London Review of Books*. URL: `http://www.lrb.co.uk/v20/n03/derek-parfit/why-anything-why-this`.

— (2011). *On What Matters*. Oxford University Press.

Peacocke, Christopher (2014). *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*. Oxford.

Prichard, H. A. (1912). "Does Moral Philosophy Rest on a Mistake?" In: *Mind* 21, pp. 21–37.

Putnam, Hilary (1987). "Truth and Convention: on Davidson's Refutation of Conceptual Relativism." In: *Dialectica* 41, pp. 69–77.

Quine, WVO (1948). "On What There Is." In: *Review of Metaphysics* 2, pp. 21–38.

— (1951a). *Mathematical Logic*. Harvard University Press.

Quine, Willard Van Orman (1951b). "Ontology and Ideology." In: *Philosophical Studies* 2, pp. 11–15.

Quinn, Warren and Phillippa Foot (1993). *Morality and Action*. Cambridge University Press.

Railton, Peter (1986). "Moral Realism." In: *The Philosophical Review* 95, pp. 163–207.

Raz, Joseph (1975). *Practical Reason and Norms*. Oxford University Press.

— (1999). *Engaging Reason: On the Theory of Value and Action*. Oxford.

Ridge, Michael (2007). "Anti-Reductionism and Supervenience." In: *Journal of Moral Philosophy* 4.3, pp. 330–348.

Risjord, Mark (2000). *Woodcutters and Witchcraft*. SUNY Press.

Rosati, Connie (2006). "Moral Motivation." In: *Stanford Encyclopedia of Philosophy* Spring 2014 edition.

Rosen, Gideon (1990). "Modal Fictionalism." In: *Mind* 99, pp. 327–354.

— (2002). "Conceivability and Possibility." In: Oxford University Press. Chap. A Study in Modal Deviance, pp. 283–308.

Ross, W. D. (2007). *The Right and the Good*. Clarendon Press.

Rovane, Carol (2013). *The Metaphysics and Ethics of Relativism*. Harvard University Press.

Ruben, David-Hillel (1987). "Explaining Contrastive Facts." In: *Analysis* 47, pp. 35–37.

Rubin, Michael (2009). "Synthetic Ethical Naturalism." PhD thesis. University of Massachusetts Amherst. URL: `http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1024&context=open_access_dissertations`.

Sandborg, David (1998). "Mathematical Explanation and the Theory of Why-Questions." In: *British Journal for the Philosophy of Science* 49.4, pp. 603–624.

Sayre-McCord, Geoffrey (1997). ""Good" on Twin Earth." In: *Philosophical Issues* 8, pp. 267–292.

— (2007). "Metaethics." In: *Stanford Encyclopedia of Philosophy* Summer 2014 Edition.

Scanlon, Tim (1998). *What We Owe to Each Other*. Harvard University Press.

— (2014). *Being Realistic about Reasons*. Oxford University Press.

Schaffer, Jonathan (2000). “Trumping Preemption.” In: *The Journal of Philosophy* 97.4, pp. 169–181.

Schilpp, P. A. (1968). *The Philosophy of G. E. Moore*. Open Court.

Schroeder, Mark (2007). *Slaves of the Passions*. Oxford University Press.

— (2010). “Value and the Right Kind of Reason.” In: *Philpapers*. URL: `http://philpapers.org/archive/SCHVAT-2`.

Schroeter, Laura and Francois Schroeter (2014). “Normative Concepts: a Connectedness Model.” In: *Philosopher’s Imprint* 14, pp. 1–13.

Searle, John (1989). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press.

Shafer-Landau, Russ (2003). *Moral Realism: A Defense*. Oxford University Press.

Shakespeare, William (2015). *Hamlet*. Electronic. MIT. URL: `http://shakespeare.mit.edu/hamlet/full.html`.

Sider, Theodore (2001). “Maximality and Intrinsic Properties.” In: *Philosophy and Phenomenological Research* 63, pp. 357–364.

— (2003). “Maximality and Microphysical Supervenience.” In: *Philosophy and Phenomenological Research* 66, pp. 139–149.

Smith, Michael (1994). *The Moral Problem*. Princeton University Press.

Soames, Scott (2004). *Reference and Description: The Case against Two Dimensionalism*. Princeton University Press.

— (2008). “Truthmakers?” In: *Philosophical Books* 49, pp. 317–327.

Sonderholm, Jorn (2008). “Why Supervenience is a Problem for Brink’s Version of Moral Realism.” In: *Journal of Philosophical Research* 33, pp. 203–213.

Strandberg, Caj (2008). “Particularism and Supervenience.” In: vol. 3. Oxford Studies in Metaethics. Oxford University Press. Chap. 6, pp. 129–158.

Street, Sharon (2008). “Oxford Studies in Metaethics.” In: vol. 3. Oxford University Press. Chap. Constructivism about Reasons, pp. 207–246.

Sturgeon, Nicholas (2009). “Oxford Studies in Metaethics.” In: vol. 4. Oxford University Press. Chap. Doubts about the Supervenience of the Evaluative, pp. 53–90.

Sverdlik, Steven (1985). “Counterexamples in Ethics.” In: *Metaphilosophy* 16.

Today, Russia (2011). “A-macing approach: US police spray peaceful students.” In: *Russia Today*. URL: https://www.rt.com/news/police-spray-peaceful-students-725/.

Varzi, Achille (2011). “On Doing Ontology without Metaphysics.” In: *Philosophical Perspectives* 25, pp. 407–423.

— (2012). “The Naming of Facts.” In: *Analysis* 72, pp. 322–323.

Varzi, Achille and Roberto Casati (1994). *Holes and Other Superficialities*. MIT.

Vayrynen, Pekka (2016). “The Supervenience Challenge to Non-Naturalism.” In: *forthcoming*. URL: http://www.personal.leeds.ac.uk/~phlpv/papers/Vayrynen-supervenience-Routledge.pdf.

Walsh, Sean Drysdale (2011). “Maximality, Duplication and Intrinsic Value.” In: *Ratio* 24, pp. 311–325.

Warnock, G. J. (1967). *Contemporary Moral Philosophy*. MacMillan.

Wedgwood, Ralph (2015). “The Pitfalls of Reasons.” In: *Philosophical Issues* 25, pp. 123–143.

Wick, Warner A. (1953). “Moral Problems, Moral Philosophy, and Metaethics: Some Further Dogmas of Empiricism.” In: *The Philosophical Review* 62, pp. 3–22.

Williams, Bernard (1981). *Moral Luck: Philosophical Papers 1973-1980*. Cambridge University Press.

— (1985). *Ethics and the Limits of Philosophy*. Taylor and Francis E-Library.

Williamson, Timothy (1996). *Vagueness*. Routledge.

— (2002). *Knowledge and its Limits*. Oxford.

— (2008). *The Philosophy of Philosophy*. Oxford.

Xenophon (2015). *Memorabilia*. Electronic. URL: http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0208%3Abook%3D4%3Achapter%3D2%3Asection%3D26.

Yasenchuk, Ken (1995). “Assimilative Moral Realism.” PhD thesis. McMaster University.

Zangwill, Nick (2006). “Contemporary Debates in Moral Theory.” In: Blackwell Publishing. Chap. Moral Epistemology and the Because Constraint, pp. 263–282.