



BIROn - Birkbeck Institutional Research Online

---

Enabling Open Access to Birkbeck's Research Degree output

In defence of robust metanormative naturalism

<https://eprints.bbk.ac.uk/id/eprint/46502/>

Version: Full Version

**Citation: Beardsley, Tobias (2021) In defence of robust metanormative naturalism. [Thesis] (Unpublished)**

© 2020 The Author(s)

---

All material available through BIROn is protected by intellectual property law, including copyright law.

Any use made of the contents should comply with the relevant law.

---

[Deposit Guide](#)  
Contact: [email](#)

**IN DEFENCE OF ROBUST METANORMATIVE  
NATURALISM**

*Tobias Beardsley*

Submitted in fulfilment of the requirements for the  
degree of Doctor of Philosophy

Department of Philosophy  
School of Social Sciences, History and Philosophy  
Birkbeck, University of London

July 2021

## Contents

<i>Abstract</i> .....	3
<i>Declaration</i> .....	4
<i>Acknowledgements</i> .....	5
<b>1. Introduction</b> .....	<b>6</b>
1.1. What is normativity? .....	9
1.2. Robust Metanormative Realism.....	13
1.3. Epistemic Naturalism .....	17
1.4. Outline of this dissertation .....	21
<b>2. Deliberative Indispensability</b> .....	<b>24</b>
2.1. Is normative scepticism self-defeating? .....	27
2.2. The argument.....	30
2.3. Deliberation and (mind-independent) normative truths .....	34
2.4. A relativistic model of (normative) deliberation .....	42
2.5. Do we have sufficient reason to engage in normative deliberation?.....	48
2.6. Conclusion .....	53
<b>3. Parfit’s Wager</b> .....	<b>54</b>
3.1. Kahane’s Wager.....	57
3.2. Normative deliberation, IBE and bootstrapping .....	60
3.3. The argument.....	66
3.4. The Normative Matrix and Normative Superdominance .....	68
3.5. From normative reasons to categorical reasons.....	79
3.6. The Deliberative Indispensability Argument revisited .....	84
3.7. Conclusion .....	87
<b>4. Ontological Implications</b> .....	<b>89</b>
4.1. Normative truths and the correspondence theory of truth.....	92
4.2. Normative truths and modal truths .....	96
4.3. Two senses of “exist” .....	99
4.4. Minimalism and fictionalism.....	104
4.5. Do normative properties have causal standing? .....	110
4.6. An appeal to supervenience .....	119
4.7. Conclusion .....	123
<b>5. Queerness</b> .....	<b>125</b>
5.1. Queerness and supervenience.....	128
5.2. Queerness and the (synthetic) <i>a priori</i> .....	130
5.3. Scepticism about the <i>a priori</i> .....	131
5.4. Queerness and irreducible normativity .....	138
5.5. Irreducible normativity and the interaction problem.....	144
5.6. Conclusion .....	152

<b>6. Triviality</b> .....	<b>155</b>
6.1. The just-too-different intuition.....	156
6.2. The Normativity Objection.....	159
6.3. The Fact Stating Argument.....	164
6.4. The Triviality Objection.....	172
6.5. Conclusion .....	181
<b>7. Normative Explanations</b> .....	<b>183</b>
7.1. Normative facts and best explanations.....	189
7.2. Reflective equilibrium .....	192
7.3. The need for intuition.....	195
7.4. Are (fundamental) normative truths conceptual truths? .....	198
7.5. Spontaneity and convergence .....	205
7.6. The problem with spontaneity .....	217
7.7. Is normativity really the best explanation for our response to agony? .....	227
7.8. Conclusion .....	237
<b>8. Reductionism</b> .....	<b>240</b>
8.1. The Reduction Argument .....	245
8.2. The problem with reductionism .....	256
8.3. Predicate D* .....	262
8.4. Is this reductionism?.....	267
8.5. Conclusion .....	271
<b>9. Conclusion</b> .....	<b>274</b>
9.1. Robust Metanormative Naturalism and the Holy Grail of Metaethics .....	275
<b>References</b> .....	<b>284</b>

## *Abstract*

There is a common perception that the very idea of a science of ethics is a hopeless chimera—that the doors of ethical truth are eternally locked to the investigative methods of science. In this dissertation, I defend a metanormative view which I believe could provide the key to a future science of ethics. According to this view, which I call “Robust Metanormative Naturalism,” there are mind-independent, irreducibly normative facts and properties, which are distinct from all descriptive facts and properties and which sometimes secure the truth of our normative judgements. But these irreducibly normative facts and properties are nonetheless perfectly continuous with science: knowledge of them can be acquired through a combination of observation and inference to the best explanation, and they are thus of the same kind as other irreducible posits of natural science, like subatomic particles and forces of nature.

In defence of this view, I develop an argument for the existence of mind-independent normative truths which is essentially a refined version of Pascal’s Wager. I argue that we have Pascalian reasons, given to us not by our contingent desires, but by the normative content of certain epistemically possible worlds, to engage in the project of normative deliberation—which is the project of seeking determinate answers to normative questions—and that this implies the existence of mind-independent normative truths. I then argue that the truthmakers for these truths are facts and properties which are irreducibly normative, rather than descriptive, but which are also involved in correct causal explanations for non-normative, empirical facts, and that fundamental normative truths can therefore be known on a genuinely empirical basis. In particular, I make the case that we cannot adequately explain why conscious beings universally respond in certain ways to certain of their conscious experiences (such as by wanting to avoid agony) without positing that these experiences have normative properties. And finally I suggest that, if we combine this (non-reductionist) metanormative view with a reductionist view of personal identity and the self, we can potentially arrive at a robust form of ethical objectivity in a fully naturalistic manner.

*Declaration*

I confirm that the work presented in this dissertation is my own and the work of other persons is appropriately acknowledged.

## *Acknowledgements*

First and foremost, I must thank my primary supervisor, Cristian Constantinescu, for his constant feedback, comments and advice throughout my entire time as a philosophy postgraduate student at Birkbeck, where I have studied under him for both an MRes and a PhD. Further thanks must go to my shadow supervisor, Hallvard Lillehammer, who provided invaluable feedback on drafts of more than one chapter of this dissertation, and in particular for his criticism of my paper, "Can Pascal's Wager Save Morality from Ockham's Razor?" which has since been published in the journal *Philosophia*. Michael Garnett also deserves thanks for discussing material with me that eventually became the basis for what is possibly the most important chapter of this dissertation.

I have presented several drafts of chapters of this dissertation at research seminars at Birkbeck, and I am grateful to all the Birkbeck students and staff with whom I have been able to discuss my work over the course of the last four years.

## 1. Introduction

Robust Metanormative Naturalism represents a synthesis of two separate views which are often regarded as in conflict. The first of these two views is Robust Metanormative Realism, similar to that defended by David Enoch. Enoch (2007) describes Robust Metanormative Realism as “the view, somewhat roughly, that there are response-independent, non-natural, irreducibly normative truths, perfectly universal and objective ones, that when successful in our normative inquiries we discover rather than create or construct” (21). The second view is Epistemic Naturalism. This is the view that all knowledge and truth, including normative knowledge and truth, if not *itself* strictly scientific in character, is at least *continuous* with the kind of knowledge and truth delivered by the natural sciences, and thus that there are no *sui generis* domains of knowledge and truth autonomous from science. Robust Metanormative Naturalism, therefore, is the view that there are response-independent, irreducibly normative truths that we have to discover, rather than create or construct, and which are continuous with, rather than autonomous from, the truths of natural science.

This view can be reduced down to three central commitments which I will defend in turn over the course of this dissertation. It is first of all committed to the existence of *categorical reasons*: reasons, whether for action or belief, which do not depend for their binding normative force on the aims or desires of agents. If we have

normative reasons to do or believe things because they are the most efficient means to the fulfilment of our ends, I believe this implies that we also have normative reasons to take the means to our ends which are not themselves dependent on these ends and which are therefore categorical in nature. Second of all, this view is importantly committed to the *causal standing* and efficacy of normative facts and properties: normative properties, just like descriptive or non-normative properties, sometimes cause us to behave in certain ways and to form certain beliefs, and they consequently make a difference in the real world through their influence on our actions and beliefs. The third commitment is to the *empirical accessibility* of *fundamental normative truths*. Fundamental normative truths are “normative truths that obtain independently of which objects have which descriptive [or non-normative] properties” (Streumer 2017, 34). And I think we can acquire knowledge of such truths, not just *a priori* through a process of rational reflection and deliberation, but empirically and naturalistically on the basis of *inference to the best explanation*.

I will argue that normative facts and properties are involved in the best explanations for certain aspects of the natural world, such as the actions and beliefs of human beings. My view is therefore closely related to the so-called “Cornell Realism” advocated by the likes of Nicholas Sturgeon and David Brink.<sup>1</sup> But I am

---

1. It is also somewhat similar to Ralph Wedgwood’s (2007) brand of metanormative realism, which likewise holds “that normative facts and properties are *causally efficacious*, and play an essential role in causal explanations of certain contingent facts” (6). Wedgwood describes his view both as a form of Platonism (3)

not claiming that normative facts and properties *directly* earn their place in our ontology by virtue of their explanatory role, but rather that it is legitimate (and indeed necessary) to appeal to these facts and properties in explaining certain aspects of the world because we already have independent, *a priori* grounds for believing in their existence and causal efficacy. In particular, I maintain that normative facts and properties are deliberatively indispensable—we are justified in believing in these facts and properties because of the role they play in deliberation—*and* explanatorily indispensable. The latter is also the main respect in which my view differs from Enoch’s non-naturalistic Robust Realism. By combining deliberative indispensability with explanatory indispensability in this way, I think Robust Metanormative Naturalism is able to avoid the problems associated with both these alternative views. It does not create a mystery about how creatures like us, in the kind of world we live in, can acquire normative knowledge and respond to normative reasons in the first place; nor does it suffer from the “Problem of Explanatory Narrowness” (Leiter 2007, 206), meaning that normative properties are too neatly tailored to only one sort of explanandum for us to think that they are real

---

and as a form of non-reductive naturalism (6): Platonism because it is based on the doctrine that “the intentional is normative” (3) (by which he means “that there is no way of explaining the nature of the various sorts of mental states that have intentional or representational content (such as beliefs, judgements, desires, decisions, and so on), without using normative terms” (2)) and non-reductive naturalism because it holds that, “even if normative facts are not *identical* to natural facts, at least all contingent normative facts are *realized* in natural facts” (6). My view is in no way based on the doctrine that the intentional is normative, however. And Wedgwood endorses an intuitionist account of normative knowledge (10), rather than the kind of best explanations account that I favour.

properties on explanatory grounds alone (211). Instead, it seeks to vindicate the human pursuit of normative knowledge and truth by coherently unifying the deliberative and explanatory perspectives.<sup>2</sup>

In the rest of this introduction, I will say more about what I mean by “normativity” and why my view involves a commitment to categorical reasons (Section 1.1). I will also explain what I mean by “realism” in the metanormative context, distinguishing between what I believe are the two main forms of metanormative realism in the process (Section 1.2). Then I will explain what I mean by “naturalism,” in particular stressing the distinction between the Epistemic Naturalism which I accept and the different, *Metaphysical* Naturalism which I do not (Section 1.3). And finally I will provide a brief chapter by chapter outline of my entire argument (Section 1.4).

### 1.1. *What is normativity?*

Normativity, as I conceptualize it, is a kind of *force* that applies specifically to the decisions and responses of conscious agents. Street (2016b) observes that, in

---

2. The deliberative and explanatory perspectives can also be called the *practical standpoint* (Street 2016c, 293) and the *theoretical standpoint* (294) respectively. Sharon Street argues that metanormative realism “brings with it a severe and irreconcilable tension between the practical and the theoretical points of view on ourselves—between our understanding of our normative judgments as true and our understanding of them as things with scientifically discoverable causal origins.” I agree that this is indeed a problem for *non-naturalistic* forms of metanormative realism, but I hope to show that Robust Metanormative Naturalism is able to harmonize the practical and theoretical standpoints without forcing us to the conclusion “that things are valuable ultimately because we take them to be” (295), which I do not regard as fully coherent.

certain circumstances, such as when a car suddenly swerves towards us on the highway, or when we see a child in pain, “we have a conscious experience of certain features of the world (the swerving car, the child’s pain) as what we can only describe as *calling for*, *counting in favour of*, *demanding*, or *requiring* certain responses on our part (evasive action, a helping response, and so on)” (126). From experiences like these, we acquire certain distinctive concepts, such as the concept of a normative reason—the concept of one thing’s *counting in favour of* or *calling for* another—in the same way that we acquire other concepts, such as *redness* or the *scent of roses*, from other conscious experiences (127). Normative concepts are irreducibly normative in the sense that they cannot be defined or explained without invoking other normative concepts, just as the intrinsic characters of certain experiences cannot be captured or described except by using these normative concepts (126). We are subject to normative force to the extent that certain features of the world call for, count in favour of, demand or require certain responses on our part, and thus to the extent that the world exemplifies these irreducibly normative concepts (as certain of our conscious experiences seem to suggest).

There are at least two kinds of normativity, which are referred to by different names in the recent literature. Jonas Olson (2011) distinguishes between *transcendent* norms, which “apply to agents categorically; their reason-giving force transcends particular aims, activities, or roles,” and *immanent* norms, which, “by contrast, are those whose reason-giving force depends on agents’ engagement in certain goal-oriented or rule-governed activities or their occupation of certain roles, such as

institutional or professional roles" (64). Brian Leiter (2015) believes that "what we call normativity is simply an artefact of the psychological properties of certain biological organisms, i.e., what they *feel* or *believe* or *desire* (or are *disposed to feel, believe, or desire*)" (66). But he draws a distinction between what we *call* normativity and "real normativity"; the latter consists of "standards of what one ought to do or believe that are not dependent for their binding force on the attitudes, feelings, or beliefs of persons" (68). Tristram McPherson (2011) distinguishes between "formal normativity" and "robust normativity": the game of chess is normative in the formal sense "simply in virtue of its being possible to play an incorrect chess move," whereas *reasons* (unlike chess) have a kind of distinctive authority that makes them *robustly* normative (232-233). And Derek Parfit (2011a) makes a similar distinction between the *reason-involving* conception of normativity, on which "normativity involves reasons or apparent reasons," and the *rule-involving* conception, on which "normativity involves requirements, or rules, that distinguish between what is correct or incorrect, or what is *allowed* and *disallowed*" (144).

When I talk about normativity, I am talking about "real" normativity (which we can also call "transcendent," "robust," or "reason-involving" normativity), as opposed to what I will call "psychological" normativity (which encompasses so-called "immanent," "formal" and "rule-involving" normativity). Real normativity involves reasons of a kind that exemplify irreducibly normative concepts: concepts linked to the conscious experience of certain features of the world calling for or counting in favour of certain responses on our part. These kinds of reasons cannot be

simply reduced to relations between our aims or desires and the efficient means towards their satisfaction. I submit, therefore, that all specifically normative reasons are either categorical reasons themselves, or ultimately stem from categorical reasons, in the sense that these normative reasons would not exist if not for certain categorical reasons. Real normativity is the kind of normativity that involves categorical reasons.<sup>3</sup>

My focus is on real normativity because it is this kind of normativity that raises substantive metaphysical and epistemological questions. It is quite obvious both that there is such a thing as psychological normativity, and that we can observe this kind of normativity in a perfectly naturalistic manner: societies have norms, rules and standards which they uphold; individuals have normative judgments and principles which they accept and endorse. And agents clearly have reasons for action and belief in the sense that they have *motives* for these actions and beliefs, and in the sense that there are various factors which *explain* these actions and beliefs. But genuine normative force is not the same thing as motivating force or explanatory power. When we talk about *normative* reasons, we are talking about reasons that

---

3. I am not making a phenomenological argument for the existence of categorical reasons here. Street reminds us that just because

we possess a certain concept [such as the concept of a normative reason], which we've come to understand in virtue of being the subject of a certain type of conscious experience, of course does not commit us to the view that there exists, "out there" in the world, some robustly mind-independent thing that "corresponds to" or is "tracked by" the concept in question. (2016, 127)

But if there is such a thing as real normativity, there must be something that corresponds to or is tracked by our irreducibly normative concepts.

actually call for or count in favour of certain responses on our part,<sup>4</sup> rather than just explaining them in a purely descriptive manner. There is a legitimate question about whether or not there really are any such reasons; and if such reasons do exist, there is also a legitimate question about how and under what circumstances we acquire knowledge of these kinds of reasons. It is these sorts of questions to which Robust Metanormative Naturalism is intended to provide answers.

### 1.2. *Robust Metanormative Realism*

I understand metanormative realism simply as the view that there are such things as normative force and categorical reasons: that certain features of the world really do call for or count in favour of certain responses on our part (and thus that our conscious experiences to this effect are not always illusory). Yet some who certainly accept the existence of normative force and categorical reasons do not actually refer to themselves as realists. Parfit believes that “reasons for acting all derive their force from the facts that give us reasons to have certain desires and aims” (2011a, 47), rather than from the desires and aims themselves (or from the fact that we have these desires or aims), but he calls his view “Non-Realist Cognitivism” (2017, 56), arguing that normative truths, while perfectly objective, “have no positive ontological implications” (2011b, 479). John Skorupski (2010) endorses a similar view he calls “cognitive irrationalism” (439), according to which normative truths are

---

4. We might also say that they give us *warrant* or *justification* for these responses.

fundamentally about “reason relations” (i.e. relations that hold between normative reasons and what they are reasons for), and according to which these reason relations are entities that are both “irreal and objective—not mind-dependent” (420). And T. M. Scanlon (2014) denies that the ontological commitments of normative truths have anything to do with what *the world* contains (24), even though he believes that many normative truths are objective in the fullest sense (94).<sup>5</sup>

To accommodate views of this kind, we need to distinguish between *Robust Realism* and what I will refer to as *minimalism*. The difference between the two, in short, is that Robust Realism applies the correspondence theory of truth to the normative domain, whereas minimalism applies the minimalist or deflationist theory. Under the correspondence theory, “[t]ruth is a matter of the intentional object of an actual or possible belief, actual or possible statement, and so on, corresponding to some real object” (Armstrong 2004, 16). The Robust Realist, therefore, construes normative truths as corresponding to real objects “out there” in the world, the real objects in question being normative facts and properties that are ontologically distinct from normative concepts and propositions (and which are the things that make normative propositions true). The minimalist, on the other hand,

---

5. Scanlon actually says that “many truths about reasons for action are both judgement-independent and choice-independent.” A subject-matter is *judgement-independent* “if it is possible for us (at least individually) to be mistaken in our judgements about that subject,” and *choice-independent* if “the standards for assessing such judgments do not depend on what we, collectively, have done, chosen, or adopted, and would not be different had we done, chosen, or adopted something else” (93-94). However, he denies any assurance that normative questions always have determinate answers, on the grounds that “the domain of practical reasons is not a unified subject matter like set theory” (104).

holds that no more substantive conditions for the truth of a normative proposition can be given other than the content of the proposition itself. For the minimalist, “to claim that it is ‘true’ that genocide is wrong is just to claim that genocide is wrong” (Wedgwood 2007, 39); and it is a normative fact that genocide is wrong only in the sense that the proposition that genocide is wrong is true.<sup>6</sup> David Copp (2017) points out that there are (at least) two senses of “fact”: the “worldly” sense, in which “facts are the *truth makers* of propositions; they explain the truth values of propositions; perhaps they are *states of affairs*” and the “propositional” sense, “according to which a fact is simply a true proposition” (34). The Robust Realist believes in normative facts in the former, “worldly” sense; the minimalist believes in normative facts only in the latter, “propositional” sense.

The views of Parfit, Skorupski and Scanlon are all versions of normative minimalism as I have described it. I reject these views in favour of Robust Realism because I think that normative truths do carry positive ontological implications. My main reason for thinking this is that normative facts and properties (and relations) genuinely seem to possess some kind of causal efficacy and standing, whereas for a minimalist like Skorupski, “[r]eason relations are unreal,” precisely because “they have no causal standing” (2010, 439). I think we sometimes respond in certain ways to features of the world which count in favour of our responding in those ways

---

6. Of course there may be more to the truth of this proposition in the sense of further reasons why genocide is wrong (such as that it causes massive suffering and death). But the point is that there is nothing to the *truth* of the proposition that genocide is wrong beyond what there is to genocide’s being wrong (according to the minimalist).

*because* these features of the world count in favour of responding in those ways, such that if those features of the world did *not* count in favour of those particular responses, we would not actually respond in those ways. If we see a child in pain, and we have a conscious experience of the child's pain as calling for a helping response on our part, and the child's pain really does call for such a helping response, and we then proceed to make this response by actually helping the child, I believe this indicates that the normative fact that the child's pain counts in favour of helping the child is one of things that *causes* us to make this response. Unlike Robust Realism, minimalism basically amounts to an implausible epiphenomenalism about the normative.<sup>7</sup> Ronald Dworkin (1996), one of the early proponents of minimalism, held that "morality and the other evaluative [i.e. normative] domains make no causal claims" (120). But I concur with Leiter in replying that "that is plainly false: the moral explanations literature from the 1980s onwards—recall Brink, Railton, Sayre-McCord, Sturgeon and others—is replete with examples of the role of causal claims in ordinary normative discourse (e.g., 'of course he betrayed them, he's an evil person')" (69).

---

7. Enoch's version of Robust Realism is not committed to the causal efficacy of normative facts (Enoch 2013, 7). But if normative facts and properties are causally efficacious, as I believe, this still entails a form of Robust Realism, because it means that these facts and properties make a difference "out there" in the real world through their effects on our responses.

### 1.3. *Epistemic Naturalism*

I believe that the role of causal claims in ordinary normative discourse, and the consequent ontological implications of (mind-independent) normative truths, strongly supports the synthesis of Robust Realism with naturalism. If normative properties make a difference in the natural world through their influence on us and our responses to them, and yet these normative properties are not themselves natural properties, this would seem to make them into *supernatural* properties.<sup>8</sup> And the constraints of parsimony very much count against accepting a commitment to the existence of robust supernatural properties.

Naturalism is sometimes defined as the view that there is no genuine knowledge outside of empirical science (BonJour 1998, 69), or that questions about what there is and what we know are reliably answered (only) by the methods of science (Leiter 2015, 64). It would indeed be fortunate if we could use scientific methods to find answers to normative questions, because these are unquestionably the most powerful and reliable methods we have for expanding our knowledge. Normative questions are among the most important questions we can ask—they are questions, after all, precisely about, or directly relevant to, what is important, or what *matters*—so we should, if at all possible, seek to find a way of leveraging our most effective methods of inquiry for the sake of answering these questions. Yet

---

8. Paul Draper (2007) defines the supernatural as that which is not part of nature but which can nonetheless affect nature (277). If normative properties have causal standing but are not natural, it follows that they must be supernatural on this definition.

unfortunately it is dubious that whatever normative knowledge we actually do possess is in any way the product of scientific methods. *Deliberation* is currently the method we use to arrive at normative conclusions. And deliberation is not the same thing as scientific experiment, especially in the moral case. We came to know that water is H<sub>2</sub>O as a result of controlled observation and experiment, but these same methods have not availed us of the knowledge that it is morally wrong to torture people for our own amusement.

I will not define normative naturalism as the thesis that normative knowledge is *necessarily* a form of scientific knowledge (although I hope it can eventually aspire to this status). Instead, I will define it as the thesis that fundamental normative knowledge can be accessed by empirical means: that fundamental normative propositions can, in principle, have empirical evidence for and against them. As Parfit says, the normative naturalist is someone who believes that “[n]ormative truths ... are like other truths about the natural world which might be empirically discovered, in the sense that some partly observable things or events might give us evidence for or against our belief in these truths” (2017, 55). I will define normative non-naturalism, in contrast, as the thesis that fundamental or *pure* normative knowledge is *never* accessible empirically, because the normative domain is completely autonomous from science. Scanlon writes that “[m]ost of the claims we commonly think of as normative are not pure normative claims, but *mixed normative claims*. They involve pure normative claims but also make or presuppose claims about natural facts” (2014, 37). The non-naturalist holds that there are pure

normative truths which do not presuppose any natural facts and which can only be known to be true *a priori*.

As a non-naturalist, Parfit denies that fundamental normative truths are ever accessible empirically. In his words, “[t]here could not be any empirical evidence for or against the belief that we have reasons to want to avoid future pain, or the belief that torturing people for our own amusement would be wrong” (2017, 58). The naturalist does not need to deny that these intuitively plausible beliefs can be known to be true (or at least reasonably believed) in the absence of empirical evidence. What I maintain is that, even if we do not need to rely on empirical evidence to know (or reasonably believe) certain normative truths, these same truths can still also be known on a genuinely empirical basis. Just because there are some synthetic *a priori* normative truths does not entail that the normative domain is completely autonomous from science. I think we learn from Quine (1976) that there is no hard and fast boundary between an *a priori* accessible domain such as pure mathematics and a branch of empirical science such as physics, because “the contrasts that people are prone to draw between pure mathematics such as arithmetic, on the one hand, and physics, on the other, can be drawn just as well between theoretical physics and experimental physics” (75). For Quine, the boundaries between domains are essentially artificial, and “[w]hen we abstract from these boundaries, we see all of science—physics, biology, economics, mathematics, logic, and the rest—as a single sprawling system, loosely connected in some portions but disconnected nowhere” (76). What ties these domains together is that they all have some kind of link to

observation: we can observe how mathematical and logical principles work in the real world (for instance, as they are used in science). The non-naturalist may insist that the normative domain has no link to observation whatsoever; but if we really do have the capacity to respond to normative reasons, the normative domain gets its link to observation through its link to our observable responses, and this remains the case even if certain normative truths can also be known *a priori*.

This Epistemic Naturalism, according to which there is no knowledge that is autonomous from science, is importantly distinct from a stronger Metaphysical Naturalism. Metaphysical Naturalism implies that there are no irreducibly normative facts and properties: to the extent that there are any normative facts,<sup>9</sup> these facts reduce to natural facts in the sense that these facts can also be stated using entirely non-normative, naturalistic concepts. If this kind of naturalism is true, then, as Parfit says, “the fact that some act is right might be the same fact as the fact that this act minimizes suffering, or is an act of which most people would approve” (2017, 57). I will classify the thesis that normative facts (such as the fact that some act is right) are the same facts as facts that are statable using entirely non-normative concepts (such as the fact that this act minimizes suffering, or is an act of which most people would approve) as normative *reductionism*. Reductionism is not the same thing as naturalism. On the one hand, reductionism is clearly compatible with supernaturalism: if normative facts can also be stated using entirely non-normative

---

9. The Metaphysical Naturalist might hold that there are simply no normative facts.

concepts, the fact that some act is right could be the same fact as a supernatural fact, like the fact that this act accords with the will of God.<sup>10</sup> And on the other hand, it may be possible to acquire (fundamental) normative knowledge *a posteriori* even though normative facts are irreducibly normative (they cannot be stated without using irreducibly normative concepts).

#### 1.4. *Outline of this dissertation*

Robust Metanormative Naturalism is precisely the view that there are irreducibly normative facts which can be known on an empirical basis. I will begin my defence of this view by providing a defence of metanormative realism more broadly in Chapters 2 and 3. Chapter 2 will assess Enoch's Argument from the Deliberative Indispensability of Irreducibly Normative truths. I argue that Enoch does show that, in at least one important sense, deliberation is committed to the existence of mind-independent normative truths. But he fails to show that deliberation is itself justified in a way that entails the existence of these mind-independent normative truths. Chapter 3 will supplement Enoch's argument with a further argument drawn from Parfit which I call "Parfit's Wager." I think that Parfit's Wager is able to show that deliberation is justified in a way that entails the existence of mind-independent normative truths. In Chapter 4, I will make the case against normative minimalism that normative truths have positive ontological

---

10. Not to imply that there are such supernatural facts, of course.

implications on account of the causal standing that normative facts and properties appear to possess. This will complete my defence of Robust Metanormative Realism.

In Chapters 5 and 6, I will consider what I believe to be the main arguments against normative non-naturalism and normative naturalism respectively. Chapter 5 will consider the Argument from Queerness against the existence of non-natural normative facts and properties. I argue that whatever force this argument has essentially reduces to the (not inconsiderable) force of parsimony and Occam's Razor. Chapter 6 responds to Enoch's and Parfit's arguments against normative naturalism, focusing in particular on Parfit's Triviality Objection, according to which certain fundamental normative claims would be rendered implausibly trivial if normative properties were also natural properties. I argue that, at best, this argument only applies to normative reductionism, not the kind of Epistemic Normative Naturalism which I advocate, and in any case, it fails to take into account properly the fact that, if certain natural facts (such as the fact that some act would maximize happiness) are identical to certain normative facts (such as the fact that some act is right), these natural facts would themselves be normative facts (albeit unobviously) by virtue of their identity with these normative facts.

In Chapters 7 and 8, I will explicitly defend (non-reductive) Epistemic Normative Naturalism. If normative properties make a difference in the real world, these properties should also be able to explain certain observable circumstances. As I stated in Section 1.1, normativity is a force that applies to our agential responses, so unless normativity is merely epiphenomenal, it must have the power to affect our

responses, and thus have the power to *explain* some of our response as well. Since normative facts and properties can influence our responses, it seems reasonable to infer that normative facts can be known on the basis that these facts are the *best explanations* for certain of our responses. This Normative Explanations Argument, I believe, is the central argument for normative naturalism. And it is this argument that I will defend in Chapter 7. In particular, I will argue that the best explanation for the empirical fact that all conscious beings have a desire to avoid agony is the normative fact that we always have a (normative) reason to want to avoid agony. Finally, in Chapter 8, I will argue against normative reductionism that normative properties are irreducibly normative. Normative properties are necessarily co-extensive with non-normative properties (there is no possible world in which it is not wrong to torture people for our own amusement, for instance), but I think this indicates, not that normative properties are identical to non-normative properties (in the sense that these normative properties can also be stated using entirely non-normative concepts), but that normative properties consist in certain *patterns* in non-normative properties, patterns which are themselves irreducibly normative.

The most plausible alternative to Robust Metanormative Naturalism, I believe, is a complete normative scepticism. And so, in the following chapter, I will begin the defence of my view by considering the sceptical alternative to metanormative realism.

## 2. Deliberative Indispensability

Normative scepticism is the view that there is no such thing as normative knowledge: that no positive normative judgement is ever known to be true.<sup>11</sup> This view is often motivated on the grounds that a belief in normative knowledge and truth saddles one with unacceptable commitments, whether metaphysical, epistemological, or both. Unlike the metanormative realist, the sceptic has no need “to explain how there can be facts that *in themselves*, that is, irrespectively of the desires, aims, roles, or activities of human beings and other agents, require, or *count in favour of*, certain forms of behaviour” (Olson 2011, 66). The realist might be obliged to posit a mysterious, inexplicable faculty of normative intuition in order to account for normative knowledge. Or he might have to make an ontological commitment to metaphysically “queer” normative entities (such as irreducibly normative favouring relations) in order to explain the normativity of normative truths. And even if the realist does not have to take on any inherently unacceptable commitments, a view which has to make room for normative knowledge and truth may just be less parsimonious than a view which manages to leave them out.<sup>12</sup>

---

11. That is, a judgement which says that something has some normative property (not just that it lacks some normative property).

12. An error theory about all normative judgements (which states that all normative judgements are systematically false because there are no normative

In this chapter, we will look at one argument *against* normative scepticism and *for* metanormative realism: Enoch's Argument from Deliberative Indispensability, according to which we are justified in believing in mind-independent normative truths because such truths are indispensable to the (indispensable) project of deciding what it makes most sense to believe and to do. Enoch (2013) argues that "objective, irreducibly normative facts are indispensable ... for deliberation, and that this indispensability suffices to justify belief in their existence" (9). For reasons I will explain, I don't think this argument ultimately succeeds, at least in the way that Enoch formulates it. But I think an important truth can be extracted from it which will help us to construct what I believe to be a successful argument for realism: the argument I call "Parfit's Wager" which I will defend in the next chapter.

Some argue that we don't actually need a direct argument against normative scepticism, because, for foundational reasons, a complete normative scepticism is completely untenable. Copp (1995) suggests that there might be no argument for metanormative realism which does not beg the question against scepticism, because any argument must involve an appeal to some norms, even if only those calling for coherence and consistency in argument. But, in Copp's view,

---

properties) has recently been defended by Bart Streumer (2017). Streumer's arguments for this view touch heavily on issues relating to the supervenience of normative properties on natural properties and the reducibility of the former to the latter, so I will not address them until Chapter 8 when I discuss normative reductionism.

this gives us no reason to accept normative skepticism. On the contrary, it suggests that an attempt to defeat normative skepticism would be an instance of the sort of enterprise Otto Neurath warned philosophers against when he urged us to remember that a sailor cannot rebuild her entire boat all at once while at sea. (46)

In what follows, I will reply to the suggestion that normative scepticism is self-defeating on a foundational level, arguing that even the most thoroughgoing normative scepticism can be motivated in such a way as to make it a serious option for the metanormative theorist (Section 2.1). The realist, therefore, does need a direct argument for realism. In Section 2.2, I will present the Deliberative Indispensability Argument in a way that makes it neutral between normative minimalism and Robust Realism.<sup>13</sup> In Section 2.3, I will argue that there is a form of deliberation, *normative* deliberation, which is committed to the existence of at least some normative truths. And in Section 2.4, I will consider and reject a relativistic model of normative deliberation, arguing that the normative domain cannot be relativistic all the way down, and thus that normative deliberation is also committed to the existence of *mind-independent* normative truths. Finally, in Section 2.5, I will examine Enoch's account of the rationality (or *intrinsic indispensability*) of normative deliberation, concluding that Enoch fails to show that we have any *normative* reason to engage in this form of deliberation, and thus that he ultimately fails to show that we are justified in accepting the commitment to (mind-independent) normative

---

13. I will postpone my treatment of this issue (minimalism vs. Robust Realism) until Chapter 4.

truths that normative deliberation carries with it. We are therefore in need of a further argument for metanormative realism.

### 2.1. *Is normative scepticism self-defeating?*

Copp argues that a complete normative scepticism cannot (or should not) be taken seriously at all: although not strictly self-refuting, normative scepticism is self-defeating and therefore, in his words, “not seriously on the agenda,” (48). Normative scepticism implies that we are not justified in believing that anything has any normative property or status, and since being epistemically justified is itself a normative status (justification, epistemic or otherwise, being a paradigmatically normative notion), if normative scepticism is true, it follows that we are not justified in believing that it is true, or that it is epistemically justified to any extent whatsoever. Copp takes this to indicate that “the skeptic is in an untenable position ... she cannot consistently accept that her scepticism has any justification” (47). Skorupski (2010), likewise, holds that: “[r]eason relations are the essence of thought, in that thought itself just is responsiveness to them. To deny that there are truths about reason relations is to deny the objectivity of thought itself” (457).<sup>14</sup> If the sceptic is committed to denying the objectivity of thought itself, it is difficult to see how her own position could be objective either. And if we already know that a

---

14. It is not quite accurate to suggest that the sceptic denies that there are truths about reasons relations: if there are no reason relations, the truth that there are no reason relations would itself be a truth about reason relations. But of course, if there are no normative truths, then there are no truths implying the existence of normative reason relations.

position cannot possibly be objective even before considering arguments in its favour, it seems reasonable enough to infer that, as Copp says, it is not seriously on the agenda.

Epistemic justification involves normative reasons for belief: considerations that *count in favour* of holding a certain belief in a way that “enhances, to an appropriate degree, the chances that the belief is *true*” (BonJour 1998, 1). In so far as the normative sceptic denies the existence of normative reasons, Copp is clearly right that she cannot consistently accept that her scepticism has any justification.<sup>15</sup> But this does not necessarily mean that there cannot be *evidence* for the truth of normative scepticism<sup>16</sup>; it does not put the sceptic “in the unfortunate position of subscribing to a theory according to which it is impossible to claim truly that there is evidence that the theory is true” (Olson 2014, 163). And even if a position is completely unjustifiable from an epistemic point of view, it might still be successfully motivated on the pragmatic grounds that it is simply “more efficacious

---

15. If normative scepticism is *false*, it actually could have some epistemic justification: false beliefs, after all, are often taken to be capable of having at least some justification. But if normative scepticism is *true*, this still implies that it could not have any justification.

16. Eliezer Yudkowsky (2007a) tells us that evidence “is an event entangled, by links of cause and effect, with whatever you want to know about. If the target of your inquiry is your shoelaces, for example, then the light entering your pupils is evidence entangled with your shoelaces.” Evidence, in at least one important sense, is a purely causal notion rather than a normative one, and thus it is not self-defeating for the normative sceptic to appeal to evidence in support of her position.

than other myths as a device for working a manageable structure into the flux of experience" (Quine 1980, 44).<sup>17</sup>

As a normative sceptic, Olson (2014) rejects the idea that normative scepticism is self-defeating by drawing a distinction between "arguments to the effect that  $p$ , or evidential support for  $p$ , and epistemic reasons to believe that  $p$ " (160). The normative sceptic is not offering arguments to the effect that there are normative reasons to believe in normative scepticism (and thus to believe that there are no normative reasons); she is offering arguments to the effect that normative scepticism is true (157). For this purpose, the sceptic can appeal to the (non-normative) notion of *indicator evidence*, according to which " $q$  is evidence that  $p$  just in case  $q$  reliably indicates that  $p$ " (162). She can use this notion, in conjunction with a principle of parsimony like Occam's Razor, to argue that scepticism is more parsimonious than realism, and that this indicates that scepticism is the more likely to be true without implying that there are normative epistemic reasons to believe either of these views.

Olson acknowledges that "appeals to Occam's Razor and considerations of theoretical simplicity seem to be appeals to *norms*" (2011, 67-68), and that this threatens to entail self-defeat for a sceptic who has to rely on these kinds of appeals

---

17. Though Quine is talking not about scepticism but about adopting alternative explanatory schemas. The relevant point is that we may have no strictly *epistemic* reasons to assign more credibility to "the myth of physical objects" than to the gods of Homer, but we may nonetheless have pragmatic reasons to do so because the former is a superior tool for predicting future experience in light of past experience (44). A theory of human beings and their behaviour of which normative scepticism is an integral part may likewise, in principle, be a more efficacious predictive tool than a theory which includes a commitment to normative facts and properties.

to argue against realism. Olson's response is to maintain that Occam's Razor is an immanent rather than a transcendent norm, and that it therefore gives us only a non-categorical reason to prefer scepticism to realism (68). In other words, Occam's Razor gives us a reason for scepticism about (real) normativity and normative reasons, not in a way that inadvertently implies the existence of such reasons, but only in the sense that, all else being equal, simpler theories are generally more conducive to the realization of our goals, including the epistemic goal of having (mostly) true beliefs. And we are surely more likely to have (mostly) true beliefs by believing one theory T over a distinct theory T\* if, all else being equal, T is theoretically simpler than T\*. Normative scepticism is not committed to the existence of categorical reasons, so it is in at least one important respect a simpler theory than metanormative realism. I think this clearly indicates that normative scepticism can neither be outright rejected as self-refuting, nor "safely ignored" as merely self-defeating (Copp 1995, 48), and that the realist therefore needs a direct argument for realism.

## 2.2. *The argument*

For the sake of providing such an argument—that is capable of blunting the edge of Occam's Razor—I will slightly modify Enoch's Argument from Deliberative Indispensability so that it is not an argument for Robust Realism as Enoch defines it (that is, as involving irreducibly normative truths that have Platonic ontological implications), but simply an argument for mind-independent normative truths (that may or may not have ontological implications of some kind). Deliberation being the

process of deciding what it makes most sense to believe and to do, the essence of this argument (as I will present it) is that deliberation necessarily involves a *commitment* to mind-independent normative truths, and that deliberation is justified, so a commitment to mind-independent normative truths is also justified.

The main argument *against* metanormative realism, I believe, is the argument that normative facts and properties are not needed to explain any aspect of the world as we are able to observe it, and that the *best explanation* for why this is the case is that normative facts and properties do not really exist. This argument first sets up an *explanatory indispensability* criterion for ontological commitment, under which belief in entities of a certain kind is justified in so far as these entities are involved in our best explanations, and then simply maintains that normative facts and properties fail to meet this criterion.<sup>18</sup> The explanatory indispensability criterion may be perfectly valid as a positive basis for believing in certain kinds of entities (such as the posits of natural science). But it is open to the realist to argue that there is another criterion under which a commitment to normative facts and properties is in fact vindicated, namely, the criterion of *deliberative indispensability*. As a criterion for ontological commitment, deliberative indispensability states that belief in a certain kind of entity is justified if that entity is necessary for deliberation. This is the basis for Enoch's defence of (Robust) metanormative realism. In his words,

---

18. Perhaps it is better to use the term "reliable" in this context, rather than a normative term like "justified." The idea is that we are more likely to have (mostly) true beliefs if we only allow ourselves to believe in the kinds of entities which are needed for our best explanations, and that this excludes a belief in normative facts and properties.

“normative truths, though not explanatorily indispensable, are nevertheless deliberately indispensable ... and this kind of indispensability is just as respectable as the more familiar explanatory kind” (50).<sup>19</sup> Indispensability arguments are used to support ontological commitments in science and mathematics, and if such arguments work in these cases, this naturally raises the possibility that an indispensability argument might work in the normative case as well.

The above outline frames the issue in explicitly ontological terms, as Enoch does (because he is arguing for Robust Realism), but if deliberative indispensability is a valid criterion for *ontological* commitment, commitment to certain kinds of *entities*, it should also be a valid criterion for commitment to certain kinds of *truths*. If we are justified in believing in certain entities, we are justified in believing in certain truths (namely, the truths that involve those entities). Yet the reverse does not automatically follow: we might be justified in believing in certain kinds of truths, but not justified in making any positive ontological commitments as a result of those truths (and Occam’s Razor may well count against doing so). For the present purposes, we can therefore take the indispensability criterion as a basis for justified

---

19. I have omitted the “irreducibly” that precedes “normative truths” in Enoch’s text, because as I am using the Deliberative Indispensability Argument, it is not an argument for *irreducibly* normative truths (just mind-independent normative truths). Streumer (2017) recognizes that this argument “may show that normative properties are indispensable to deliberation. But it does not show that if normative properties were identical to descriptive properties, this would undermine our reason to engage in deliberation” (23).

belief in (mind-independent) normative truths, while remaining neutral about the exact ontological implications of these truths.<sup>20</sup>

What does it mean to say that specifically mind-independent normative truths are indispensable to deliberation, or that deliberation involves a commitment to mind-independent normative truths? The basic idea is that the non-existence of such truths would completely undermine our reason to engage in deliberation. If what we are essentially doing when we deliberate is trying to find determinate answers to normative questions, then, if there were no (mind-independent) normative truths for us to discover, deliberation would always inevitably fail, and thus, on Enoch's terms, we would have no reason (or at least, not sufficient reason) to engage in deliberation at all. Therefore, if we *do* have sufficient reason to engage in deliberation, (mind-independent) normative truths must exist: because deliberation is fundamentally impossible without them, either such truths *don't* exist and we *don't* have sufficient reason to engage in deliberation, or we *do* have such reason and normative truths *do* therefore exist.

With this in mind, we can see a simple way of presenting the Deliberative Indispensability Argument for metanormative realism:

---

20. In one sense, if a commitment to normative truths is justified, then a commitment to the *existence* of such truths is justified. So it seems that, even if we try to deny that normative truths have ontological implications of some kind, we are still making an ontological claim. But Parfit and Skorupski both maintain that we have to distinguish between two senses of "exist": an ontological and a non-ontological sense, and that *truths* (normative or otherwise) only "exist" in the non-ontological sense. We will return to this issue in Chapter 4 (Section 4.3).

- (1) If there were no (mind-independent) normative truths, we would not have sufficient reason to engage in deliberation.
- (2) We *do* have sufficient reason to engage in deliberation.
- (3) Therefore, there are (mind-independent) normative truths.<sup>21</sup>

This formulation captures the notion of deliberative indispensability (premise (1)), but it puts the point in terms of the dependence of the reason (or justification) we have for deliberation on a certain kind of truths, (mind-independent) normative truths, rather than in terms of any explicitly ontological commitments that deliberation is supposed to license. This may seem to be a problem: isn't deliberative indispensability modelled on explanatory indispensability in the philosophy of science and mathematics? As an argument for Robust Realism, the Deliberative Indispensability Argument does rely on an analogy with explanatory ontological commitments, such as the commitment to electrons (54-55). But as an argument just for realism, it will not lose plausibility points by leaving open more ontological questions rather than fewer.

### 2.3. *Deliberation and (mind-independent) normative truths*

Having established that we can state the Deliberative Indispensability Argument in adequately non-ontological terms,<sup>22</sup> let's take the premises of the

---

21. This argument can also be stated in terms of deliberation being *warranted* or *justified*.

22. I am not ruling out the possibility that a commitment to normative truths either goes hand in hand with an ontological commitment or is a fully ontological commitment in its own right. I just mean that the argument, as I have presented it, leaves the ontological question as open as possible while still capturing all that is

argument one at a time. Start with premise (1): why do there need to be (mind-independent) normative truths for us to have sufficient reason to engage in deliberation? In defence of the idea that deliberation is committed to (mind-independent normative truths, Enoch appeals to the *phenomenology* of deliberation. He describes deliberation as “an attempt to eliminate arbitrariness by discovering (normative) reasons,” which is “impossible in a believed absence of such reasons to be discovered,” and because deliberation “feels like trying to make the *right* choice,” and the phenomenology of deliberation is therefore similar “to that of trying to find an answer to a straightforwardly factual question” (72-74).<sup>23</sup> It is no doubt true that we often do ask ourselves questions about what we ought or have reason to do with the aim of finding the *right* answer, on the assumption that such an answer already exists independently of us. Normative questions are clearly a *part* of deliberation, and normative truths are obviously indispensable for success in answering questions of this kind. But does that indicate that deliberation as a whole involves a *necessary* commitment to specifically *mind-independent* normative truths? Perhaps only *some* of the questions we ask ourselves in deliberation call for determinate normative

---

needed in the deliberative indispensability criterion (by linking deliberation and our reason to engage in it with (mind-independent) normative truths).

23. In the “worldly” sense of “fact,” a factual question is one which demands some kind of ontological answer. In the “propositional” sense, a factual question is simply one that demands a true answer (perhaps under a non-realist theory of truth). I take a “straightforwardly factual question,” in this context, to mean a question that demands a determinate answer: one that we have to discover, rather than create.

answers. If so, the unavailability of such answers might not make deliberation into a completely futile endeavour.

Olson points out that “the question one is trying to answer in deliberation is often what one most wants (to do), or most desires (to do)” (2014, 173). If deliberation does not necessarily involve asking oneself normative questions like what one ought to do, or what it makes most sense for one to do (what it makes most sense to do being what there is most normative reason to do), then it does not necessarily involve a commitment to (mind-independent) normative truths. If it is plain enough that we often ask ourselves what we ought to do (and that in asking this question, we are aiming at a determinate answer), it is surely equally plain that we often ask ourselves what we most want (to do), and that we are aiming at equally determinate answers when we do so. If asking ourselves what we most want (to do) counts as genuine deliberation, it would seem that we can deliberate successfully in the complete absence of any normative truths, mind-independent or otherwise: all we need are psychological truths about our desires and empirical truths about how to realize them effectively. This would seem to refute premise (1) of the Deliberative Indispensability Argument: we can deliberate successfully without (mind-independent) normative truths, because deliberative questions are often not normative questions at all, and consequently, the non-existence of normative truths would not necessarily undermine whatever reason we have to engage in deliberation in the first place.

But suppose that there are no (mind-independent) normative truths, and hence that we can only deliberate successfully by asking ourselves what we most want (to do), rather than what we ought to do (or what it makes most sense for us to do in the sense that involves normative reasons). In this scenario, deliberation would still be akin to answering straightforwardly factual questions, and we would often find determinate answers to the questions we ask ourselves in deliberation. Would this refute premise (1)? Just because deliberation (of a kind) can succeed without (mind-independent) normative truths does not mean that we would still have sufficient reason to engage in deliberation in the absence of such truths. If there are no (mind-independent) normative truths, we might have no reason at all to ask ourselves what we most want (to do). We certainly do not have *normative* reason to ask ourselves what we most want (to do) if there are no normative truths. Olson would say that we have reason to deliberate (in the absence of normative truths) in the sense that we have non-normative, “hypothetical” reasons to do so, and that “hypothetical reasons claims are true only if they reduce to empirical claims about agents’ desires and (actual or believed) efficient means of bringing about the satisfaction of these desires” (153). In this sense, we clearly do have sufficient reason to ask ourselves what we most want (to do), simply as an efficient means of bringing about the satisfaction of our desires (the claim to this effect being a hypothetical reason claim that the normative sceptic can accept consistently with his view).

In defence of premise (1), we could deny that there are any (non-normative) “hypothetical” reasons of the kind Olson appeals to (and thus that, if we have any

reason to engage in deliberation, the “reason” in question must be normative), or we could deny that simply asking ourselves what we most want (to do) constitutes genuine deliberation. The first option is not promising: as Olson points out, the term “[r]eason” is notoriously ambiguous and there is clearly a sense of the term that fits the proposed understanding of hypothetical reasons” (154).<sup>24</sup> We cannot answer the sceptic’s objections just by insisting that an ambiguous term like “reason” only has one usage.<sup>25</sup> Enoch himself would probably take the second option: he distinguishes between the two different activities of trying to make up our minds and trying to make the decision it makes most sense to make (2013, 77). It is only the latter (according to Enoch) that involves a commitment to (mind-independent) normative truths, and it is *this* activity, the activity of deciding what it makes most sense to do, that he understands by “deliberation” (77-78). Olson, on the other hand, seems to take deliberation to be the former activity, the activity of trying to make up our minds. Enoch and Olson disagree on whether or not deliberation is necessarily committed to (mind-independent) normative truths, it would appear, largely because they understand “deliberation” differently.

---

24. In support of this claim, Olson gives as an example the fact “that we might say there was reason for Hitler to invade Britain during the Second World War,” meaning “only that Hitler had some desire (e.g., a desire to win the war) that would likely have been satisfied, had he invaded Britain” (154). This does not establish that hypothetical reasons claims are never normative, but I think it does show that there is a non-normative “hypothetical” use of the ambiguous term “reason.” We will return to this example in the next chapter (Section 3.5).

25. The relevant usage of “reason” (for the purposes of the realist) being the one that means “a fact or consideration that *counts in favour* of something” (154).

“Deliberation,” like “reason,” is an ambiguous term; we should not insist that the only correct usage is the one that involves a necessary commitment to normative truths. In the same way that reasons can be either normative or non-normative, there are clearly normative and non-normative forms of deliberation. And just as the non-normative form of deliberation (the activity of trying to make up our minds or of trying to decide what we most want) presumably involves a necessary commitment to non-normative reasons (hypothetical reasons, in Olson’s terminology), the *normative* form of deliberation (the activity of trying to decide what we ought to do or what it makes most sense for us to do) involves a necessary commitment to normative reasons,<sup>26</sup> and hence to normative truths. This is the kind of deliberation (let’s call it “normative deliberation”) that the Deliberative Indispensability Argument must (of course) appeal to. And if there are no normative truths, we clearly have no normative reason to engage in normative deliberation.<sup>27</sup> But could we still have sufficient normative reason to engage in normative deliberation if there were no *mind-independent* normative truths?

---

26. Olson responds to Enoch in part by denying that the decisions we make through non-normative, purely desire-based deliberation are arbitrary: “[i]t might ultimately be arbitrary what desires I in fact have, but given that I do have those desires there need be nothing arbitrary about my choice” (174). This non-arbitrariness seems to suggest that non-normative deliberation is committed to non-normative, hypothetical reasons in the same way that normative deliberation is committed to normative, categorical reasons.

27. Though we might still have some hypothetical reason to engage in normative deliberation, since doing so would probably serve some of our desires.

The evidence of what it *feels* like to engage in (normative) deliberation cannot be used to confirm the *existence* of mind-independent normative truths (the phenomenology of deliberation is hardly a reliable guide for that—it could easily feel like we are discovering mind-independent normative truths in deliberation even though we are doing no such thing),<sup>28</sup> but that does not mean that it cannot be used to confirm the *commitments* of (normative) deliberation, whether or not there is anything in reality that corresponds to these commitments. Introspective evidence probably can tell us (though not infallibly) what we are fundamentally committed to when we ask ourselves what it makes most sense to do. If (normative) deliberation genuinely *feels* like an attempt to discover, rather than create, determinate answers to normative questions, just as scientific inquiry feels like trying to discover (not create) determinate answers to scientific questions, then we have good evidence that it genuinely has this goal (the goal of finding determinate answers to normative questions), and thus that, in the absence of such answers in the form of mind-independent normative truths, deliberation of the normative kind is inevitably doomed to fail.

---

28. At the end of his discussion of the Deliberative Indispensability Argument, Olson concludes that it “establishes at most that error theorists are committed to holding that a certain kind of deliberation [what we are calling “normative deliberation”] ... is illusory” (2014, 177). Premise (1) of our formulation of the argument does not imply that (normative) deliberation is not illusory; that is the job of premise (2) (which we will address in Section 2.5). So the phenomenology of (normative) deliberation can still be good evidence for the truth of premise (1) (though probably not for the truth of premise (2)).

Normative deliberation clearly does involve trying to arrive at an independent answer—independent, that is, of the particular deliberating agent. When we deliberate in a normative manner, we are *looking* for answers to our questions and not trying to invent those answers; we would probably not even bother trying to answer normative questions at all if we thought we could just make the answers up.<sup>29</sup> If we are intentionally making up the answers, we are most likely just trying to make up our minds, and hence not really asking ourselves what we ought to do or what it makes most sense (for us) to do. Not to imply that if we *are* trying to make up our minds, we are therefore intentionally making up answers: when we ask ourselves what we most want, we are looking for a determinate answer to this question. And if there is an element of independence involved in asking ourselves what we most want, there is surely a similar element involved when we ask ourselves what it makes most sense to do.<sup>30</sup>

---

29. Taking Enoch's example, if I am not looking for a determinate answer, I might ask myself *whether* to join a law firm or apply to graduate school in philosophy, but I would not ask myself whether I *ought* to join a law firm or apply to graduate school in philosophy (or which of these things it *makes most sense* for me to do) (2013, 72). If I *said* that I was asking the second question, but also said that the answer to this question was my decision to make, this would tend to indicate that I am really asking the first question and have simply confused the two. If we accept that we can ask both questions, and that these questions are indeed distinct, we should also accept that normative deliberation strives at independent answers (independent at least of the particular deliberating agent).

30. There is a sense in which facts about our desires are *mind-dependent*: if there were no minds, there would be no facts about our desires. But facts about our desires are facts about things inside our skulls, and what is going on inside our skulls is just as much a matter of fact as what is going on in the universe outside our skulls.

Yet just because there is an *element* of independence involved does not mean that normative deliberation requires normative truths that are genuinely mind-independent. It could be that, while the phenomenology of (normative) deliberation clearly indicates that it seeks answers that are not created by the individual deliberating agent, these answers do not have to be independent of a normative framework that is entirely created by us. As a deliberator, the answers I am looking for may not be answers that *I* am creating, but they could nonetheless be answers that have already been created by others before me (my society, say). I might not be at all *aware* of this, however, and since the framework I am consulting may be almost completely independent of *me* as an individual agent, I will have the distinct impression of trying to answer a straightforwardly factual question. This seems to suggest that we can account for the phenomenology of (normative) deliberation without introducing a commitment to normative truths that are ultimately independent of a (mind-dependent) framework created by us.<sup>31</sup>

#### 2.4. *A relativistic model of (normative) deliberation*

Deliberation might have more than one “level”: on the first level, we decide what we most want, on the second level, we create a normative framework in order

---

31. Against this, one could argue that the phenomenology of deliberation would simply not be the same in a relativistic scenario; we would ask ourselves “what does such-and-such a framework tell us to do?” rather than “what does it make most sense to do?” But this assumes that we would be conscious of the fact that we are consulting a (mind-dependent) framework if indeed that is what we are doing in normative deliberation.

to help us achieve what we want, and on the third level, we decide what it makes most sense for us to do relative to the framework we have created. The first and third levels are equivalent to the non-normative and normative forms of deliberation (respectively), both of which feel like trying to find answers to factual questions. The second, on this model of deliberation, is an intermediate, “bridge” form that generates normative deliberation out of non-normative deliberation. It is not akin to finding answers to factual questions, but the phenomenological evidence is not strong enough to show that deliberation has to aim at mind-independent answers all the way down. The metanormative realist and the sceptic should both accept that there are two forms of deliberation (non-normative and normative) which *feel* like they require mind-independent answers, but if there is a third form of deliberation bridging the gap between non-normative and normative deliberation, and on which normative deliberation actually depends, then in fact only the first, non-normative form of deliberation (asking ourselves what we most want) aims or needs to aim at fully mind-independent answers. In this scenario, the best theory of truth for normative truths would seem to be a coherence theory,<sup>32</sup> in which case, metanormative realism (Robust or otherwise) would be false.

---

32. Quine (1979) would agree. In his view, “[s]cience, thanks to its link with observation, retains some title to a correspondence theory of truth; but a coherence theory of truth is evidently the lot of ethics” (477). The coherence theory “involves the rejection of realism about truth,” and hence the rejection of “the principle of bivalence (according to which every proposition is either true or false) and the principle of transcendence (which says that a proposition may be true even though it cannot be known to be true)” (Young 2018). Applied to normative truths, the coherence theory involves rejecting the mind-independence of such truths.

How would deliberation work under this model? Olson presents us with an instructive analogy:

consider (regulative) rules of football. It is in a sense arbitrary what these rules are, but it is not arbitrary for referees to make decisions based on them. For example, it is in a sense arbitrary that goalkeepers are not allowed to pick up the ball by hand when it has been played by a team-mate ... But given that this is now a rule it is not arbitrary for referees to award a free kick to the opponent team when this happens. (2014, 174)

The rules of football are a framework we create, presumably after a process of non-normative deliberation (asking ourselves what we most want). This enables referees to ask themselves what decisions they ought to make (or what decisions it makes most sense for them to make) and come up with non-arbitrary answers that are independent of them as individuals, but not genuinely mind-independent, by consulting the relevant framework. This model of deliberation preserves the commitment normative deliberation clearly has to normative truth, thus accounting for the phenomenology of deliberation. It also has the advantage of unifying deliberation by linking the normative form with the non-normative form (via the creation of normative frameworks) and explaining how the former is actually derived from the latter. And it does this without having to endorse a commitment to categorical reasons and mind-independent normative truths.

But if normative deliberation (as opposed to non-normative deliberation) is completely relativistic in nature, can it really arrive at the non-arbitrary normative answers it appears to seek? If it is not arbitrary for a football referee to award a free kick in a particular situation, and this is so because of the current rules of football,

then it must be non-arbitrary for him to base his decision on the current rules of football (rather than, say, the old rules of football or even the rules of rugby). He must, therefore, have sufficient reason to use that particular framework rather than another—and surely he does have sufficient reason to base his decision only on the current rules of football. If the reason he has is entirely non-normative, then his use of that framework will be normatively arbitrary, in which case the decision he arrives at on the basis of that framework will also be normatively arbitrary. If this is how deliberation works, then normative deliberation cannot arrive at answers that are normatively non-arbitrary. On the other hand, if he has sufficient normative reason to base his decision on the current rules of football, then his decision will not be normatively arbitrary, but this non-arbitrariness would rely on a normative truth that is not relative to the framework on which he is basing his decision: the normative truth that he has sufficient normative reason to base his decision on the current rules of football (rather than on some alternative set of rules). Could this normative truth be relative to some other framework? There might be higher-level normative frameworks which decide between lower-level frameworks (like the current rules of football). And if one has sufficient normative reason to base a decision on a certain lower-level framework, this normative truth could be relative to a higher-level framework. But then one would need sufficient normative reason to use that particular higher-level framework, thereby presupposing another normative truth that is not relative to the higher-level framework. No matter how many levels of frameworks one goes through, at each level, there will always be a need for a

normative truth that is not relative to the framework at that level, unless, that is, one's first-order decision is to be normatively arbitrary.

If normative frameworks are things that we create, there cannot be infinitely many of them, so either there are some normative truths that are not relative to frameworks we create, or normative deliberation cannot arrive at normatively non-arbitrary answers. In other words, if one has normative reason to make a particular decision because it conforms to a certain framework, one must also have normative reason to follow that framework, and if one has such reason because following that framework itself conforms to some higher-order framework, then one must have normative reason to follow *that* framework, or else one's decision is going to be normatively arbitrary. And since we cannot create an infinite series of frameworks, not all normative reasons can come from frameworks that we create. If all normative truths and normative reasons are *mind-dependent*, therefore, all our decisions are normatively arbitrary, because the frameworks on which we base these decisions are themselves normatively arbitrary. That does not mean that our decisions would be arbitrary in some non-normative sense, such as the sense in which a decision is not arbitrary if there is sufficient hypothetical reason for it; the normative domain could be relativistic all the way down without rendering all of our decisions *completely* arbitrary. But we are asking whether or not normative deliberation is a viable project in the absence of mind-independent normative truths, and since the absence of such truths would render our decisions *normatively* arbitrary, and the *normative* form of

deliberation clearly aims at *normatively* non-arbitrary answers, it certainly seems that the answer must be no.

One might argue that premise (1) of the Deliberative Indispensability Argument is still false, because we do have sufficient *non-normative* reason to engage in normative deliberation even if it cannot achieve its aim of discovering mind-independent normative truths. Olson insists that the normative error theory does not discredit normative ethics because “[t]he human predicament is such that we need to ‘find principles of equity and ways of making and keeping agreements without which we cannot hold together’,” and “[o]ur means of achieving this is to engage in first-order normative theorizing” (198). This could well be true, but if we understand premise (1) as the claim that we would not have sufficient *normative* reason to engage in normative deliberation if there were no mind-independent normative truths, then it is not touched by the possibility that we might have sufficient non-normative reason to engage in this project even if there were no such truths. This, therefore, is how premise (1) must be understood: we could not have sufficient normative reason to engage in normative deliberation unless there are mind-independent normative truths. If the normative domain was relativistic all the way down, all normative truths would be relative to frameworks that are normatively arbitrary, and normative deliberation would not be able to arrive at normatively non-arbitrary answers. Though we might still have non-normative reason to seek out such answers, we would not have any *normative* reason to do so.

2.5. *Do we have sufficient reason to engage in normative deliberation?*

In so far as premise (1) states that we would not have normative reason to engage in normative deliberation without mind-independent normative truths, it is indeed true. Now for premise (2) of the Deliberative Indispensability Argument. In Olson's sense, we clearly have sufficient non-normative (hypothetical) reason to engage in non-normative deliberation (asking ourselves what we most want). But to generate the conclusion that there are mind-independent normative truths, we must have sufficient *normative* reason to engage in *normative* deliberation.

What we need is an account of the rationality of normative deliberation, and such an account is what Enoch attempts to supply. According to Enoch, there are some projects we engage in which are what he calls "intrinsically indispensable." A project is intrinsically indispensable if it is "rationally non-optional" in the sense that disengaging from this project is not a rationally acceptable option (2013, 70). One of these projects is the reasoning project—the project of using reason, in the form of logic and rules of inference, to expand our knowledge; another is "the project of trying to find out what is going on in the world outside our minds" (64). These are both projects which, in Enoch's view, we cannot rationally opt out of; and if we cannot rationally opt out of a project, we must have sufficient reason to engage in it. But these are not the only intrinsically indispensable (rationally non-optional) projects: the explanatory project, which is the project we are engaged in when we try to understand the world through science, is in this same sense intrinsically indispensable, as is, most importantly for the metanormative realist, the deliberative

project (more specifically, the project of normative deliberation: trying to decide what it makes most sense to believe and to do).

In defence of the idea that the reasoning, explanatory and deliberative projects are rationally non-optional, Enoch makes an appeal to human nature, claiming that “[w]e are reasoning creatures” (64), that “[w]e are explaining, understanding creatures” (60), and that “we are *essentially* deliberative creatures” (70). Our nature may be such that we *cannot* disengage from these projects, not in the long term at least, but, as Enoch acknowledges, it is not clear that this suffices to make these projects *rationally* non-optional (62). In fact, if we simply cannot disengage from these projects because of our nature, this could indicate that they are not rationally non-optional at all, because for that very reason, they are not rational (although they are therefore not irrational either). Richard Joyce (2001) maintains that “rationality must be something we can fall short of ... If we are bound to follow a principle [or a project, in this case] as a matter of psychological fact [or because of our nature], then we cannot be considered *rational* to do so” (57). If Joyce is right, then if we cannot but engage in these kinds of projects, it is not *rational* for us to do so, and it could not be irrational for us to opt out of doing so, and thus these projects are not actually intrinsically indispensable in Enoch’s sense.

Enoch insists that even if we can opt out of these projects, it just seems *prima facie* irrational for us to do so; he takes the explanatory project to be a project from which “even if we *can*, it seems we *should* not so disengage” (2013, 60). He also calls this project “one of tremendous importance for us” and speaks of rationally non-

optional projects as being “valuable” (60-61). Enoch himself does not make this clear, but I suggest that the rational non-optionalness of a project (on Enoch’s account) is the result of two conditions: 1) the project is of very high value or importance for us (in some sense), and 2) the project has the high value or importance for us that it has because of some inescapable part of our nature. If a project is of enough value or importance for us, and this value or importance is generated, in large part, by our human nature, such that we cannot somehow escape the high value or importance this project has for us (except perhaps by completely changing our nature), then we are inescapably irrational if we nonetheless try to opt out of this project. *This* is arguably what it means for a project to be rationally non-optional (and hence intrinsically indispensable).

But what makes a project especially valuable or important for us?<sup>33</sup> Again, Enoch doesn’t specify, but I think we can safely assume that a project is of value or importance *for us* to the extent that engaging in this project helps us to realize our values.<sup>34</sup> This allows us to say that, if the realization of our values depends to a great enough extent on our engaging in a particular project, and we are prevented from

---

33. Enoch calls the explanatory project “one of tremendous importance *for us*” (my emphasis), not just “tremendously important.” I don’t think the kind of importance required for rational non-optionalness has to be absolute or intrinsic importance: relative importance should be enough.

34. I will remain neutral on what exactly a value is here (whether they are things we desire to desire after some process of idealization, for instance, or whether they are things we *ought* or have *reason* to desire): what matters for the argument is that a project is valuable to us to the extent that it helps us realize our values, whatever “our values” might be.

escaping this dependency by our nature, then this project is rationally non-optional, so long, that is, as it would be irrational for us to opt out of a project that we heavily depend on, because of our nature, for the realization of our values. And there clearly is at least one sense in which it would indeed be irrational for us to opt out of such a project: in the “hypothetical” sense of “reason,” we have every reason not to opt out of a project that is crucial to the satisfaction of our desires, and the satisfaction of our desires is bound to be heavily linked to the realization of our values. Therefore, if the realization of our values depends on our engaging in a certain project, and our nature prevents us from escaping this dependence (because we cannot change our values, for instance), then we are inescapably irrational, at least in the non-normative, hypothetical sense, if we try to opt out of this project.

It is probably true that our nature determines us to have certain values rather than others, and that we depend on certain projects (like the reasoning, explanatory and deliberative projects) to help us realize our values. And as such, it is also probably true that we are irrational in the hypothetical sense if we try to opt out of these projects, and thus that these projects are intrinsically indispensable in at least one important sense. But if we only have non-normative reason to engage in intrinsically indispensable projects, then even if the deliberative project is intrinsically indispensable, this will not entail the existence of mind-independent normative truths. As we have already observed, we might have sufficient non-normative reason to engage in normative deliberation in the absence of any normative truths at all (mind-independent or otherwise). If normative scepticism is

true, we would still have at least some hypothetical reason to engage in first-order normative inquiry. And if normative deliberation really is a rationally non-optional project in the sense that it is indispensable to the realization of our values, then we surely would have sufficient hypothetical reason to engage in this project in the absence of (mind-independent) normative truths. If normative deliberation is rationally non-optional for the reasons Enoch suggests, its being rationally non-optional does not in fact justify the conclusion that there are mind-independent normative truths.

But since we do have sufficient hypothetical reason to engage in normative deliberation, premise (2)—that we have sufficient reason to engage in deliberation—is still true, and as we have established, premise (1) is also true: normative deliberation does involve a commitment to mind-independent normative truths. So why doesn't the conclusion—that there are mind-independent normative truths—follow? As we have formulated the Deliberative Indispensability Argument, the move from premises (1) and (2) to the conclusion rests on an equivocation: premise (1) uses the normative sense of "reason" whereas premise (2) uses the non-normative, "hypothetical" sense, and this is why the conclusion doesn't follow. For premise (1) to be true, it has to mean that we would not have sufficient *normative* reason to engage in (normative) deliberation without mind-independent normative truths. And for all Enoch establishes, premise (2) is only true in the sense that we have sufficient *non-normative* reason to engage in (normative) deliberation. So

premise (2) does not trigger the antecedent of premise (1), and the conclusion therefore does not follow.

## 2.6. *Conclusion*

We must conclude that, as it stands, Enoch's Argument from Deliberative Indispensability fails. The normative sceptic can accept both that normative deliberation involves a commitment to mind-independent normative truths and that normative deliberation is intrinsically indispensable, because the combination of these ideas does not have to imply that a commitment to mind-independent normative truths is justified. The sceptic can reply that deliberation, normative or otherwise, is only rationally non-optional in the sense that we have sufficient hypothetical reason to engage in it. And if (normative) deliberation can be rationally non-optional without implying the existence of mind-independent normative truths, then even in the absence of such truths, we would still have sufficient reason to engage in normative deliberation. If metanormative realism is to be vindicated on broadly deliberative grounds, therefore, we are going to need a different account of the rationality of normative deliberation. This is what I will attempt to provide in the next chapter.

### 3. Parfit's Wager

Is there a sound argument for the conclusion that we not only have sufficient non-normative reason to engage in normative deliberation, but sufficient *normative* reason as well? If there is, we could defend premise (2) of the Deliberative Indispensability Argument in a way that *does* entail the existence of (mind-independent) normative truths, without relying on an equivocation between the normative and non-normative senses of "reason." I believe there is such an argument.

In a particular passage from *On What Matters: Volume Two* (a passage that, as far as I am aware, has yet to receive adequate attention in the literature), Parfit uses what is essentially a refined version of Pascal's Wager, which I will call *Parfit's Wager*, to argue for the existence of irreducibly normative truths. I think this argument shows that we can use the specifically normative form of deliberation to arrive at the conclusion that there are some normative reasons and truths, and thus, consequently, the conclusion that the normative form of deliberation is (epistemically) justified. If we can use normative deliberation itself to arrive at the conclusion that normative deliberation is justified, I believe this shows that we do have, not only sufficient non-normative or "hypothetical" reason, but sufficient

*normative* reason to engage in this form of deliberation.<sup>35</sup> And when combined with premise (1) of the Deliberative Indispensability Argument, this entails the conclusion that there are mind-independent normative truths.

The gist of this argument is that, since normative reasons are *epistemically possible*—we do not know with anything approaching certainty that there are no such reasons—if we believe in normative reasons, we might be believing what we ought to believe, and we are in fact more likely to be believing what we ought to believe, and less likely either to be believing what we ought not to believe, or to be failing to believe what we ought to believe, compared to not believing in normative reasons (whether we actually believe that normative reasons don't exist, or simply suspend judgment on the matter). For this very reason, we must have *at least some* normative reason to believe in normative reasons, and this reason that we have straightforwardly implies the existence of normative reasons. We can also add that, since this meta-reason to believe in normative reasons in no way derives from our aims or desires (Parfit's Wager does not itself rest on an appeal to our aims or desires), this reason must be a categorical reason, and thus that at least some of our reasons, contrary to any form of normative scepticism or non-realism, are normative and categorical in nature.<sup>36</sup>

---

35. In the same way that we can use *non-normative* deliberation to arrive at the conclusion that we have sufficient *non-normative* reason to engage in non-normative deliberation.

36. To be clear, two separate claims are being made here: first, that Parfit's Wager establishes the existence of at least some normative reasons (the normative

Guy Kahane (2017) has recently made a similar Pascalian argument against “nihilism,” by which he means “both the view that nothing has final value, and that there are no reasons to want, do or feel anything” (331). In Section 3.1, I will explain why this argument doesn’t quite succeed as an argument against normative (or evaluative) scepticism, and suggest why Parfit’s Wager can succeed where “Kahane’s Wager” fails. In Section 3.2, I will prepare the ground for the argument by stressing the analogy between normative deliberation and the form of reasoning known as inference to the best explanation (IBE), in terms of the need they both seem to have to pull themselves up by their own bootstraps from an epistemic point of view. Then, in Sections 3.3 and 3.4, I will present the Parfit’s Wager argument for the existence of normative reasons and truths, in the process supplying and defending an additional premise (a principle which I call “Normative Superdominance”) that I believe the argument needs in order for the conclusion to follow. In Section 3.5, I will make the argument that, given what normative reasons are, and given the way in which the existence of these kinds of reasons can be established using Parfit’s Wager, at least some of these reasons must be categorical. And finally I will explain, in Section 3.6, how the kind of Pascalian reasoning involved in Parfit’s Wager actually vindicates the Deliberative Indispensability Argument and establishes the existence of mind-independent normative truths.

---

reason we have to believe in normative reasons itself being one such reason); and second, that, given the *way* in which Parfit’s Wager establishes the existence of normative reasons, at least some of these reasons must be categorical.

### 3.1. Kahane's Wager

Kahane argues that we have powerful pragmatic reasons to believe that normative or evaluative nihilism is false, and that, in contrast, some things really do matter (347). He explains that this is because,

If nothing matters, this [the fact that nothing matters<sup>37</sup>] doesn't matter either. But if nothing matters, and we believe that, then—although it won't matter *whether* anything would still matter to us—it's likely that far fewer things *would* matter to us. If nothing matters then this result of belief in nihilism of course also won't matter. But it *would* matter, and matter greatly, if we falsely believe in nihilism and stop, in this way, to care about the things that do matter. (347-348)

If some things really matter, and yet we believe that nothing matters, this would clearly be a bad thing: we would stop caring about things, such as suffering, which (in this scenario) really matter, and which we therefore very much ought to care about. However, if we instead believe that some things matter, and yet nihilism is in fact true and nothing really matters at all, no negative consequences will follow, precisely because nothing matters. Moreover, if some things do matter, then, so long as our evaluative beliefs are roughly on track (we are not “deeply mistaken in thinking that suffering is bad, justice is good, and so forth” (337)), believing that nihilism is false and that some things matter will have good consequences, because it will lead us to care about many things that (in this scenario) we actually ought to care about. And if nihilism is in fact true and nothing really matters, if we believe (albeit correctly) that nihilism is true and that nothing matters, there will be nothing

---

37. Or correctly believing that nothing matters (for that matter).

good about this because, once again, nothing at all would matter. This seems to suggest that we ought to believe, or try to believe, that nihilism is false and that some things really do matter, because of the potential gains involved in doing so and the potential risks involved in not doing so, combined with the complete lack of potential gains involved in the belief in nihilism and the lack of potential risks involved in the belief that some things (really) matter.

I think Kahane is broadly on the right track, but his argument relies on an assumption about our normative or evaluative psychology which I just don't think we need to make in order to construct a successful wager argument against nihilism or normative/evaluative scepticism. Kahane's Wager is based on a principle he calls "Belief Loss," according to which "[c]oming to believe in nihilism will result in our coming to lose our substantive evaluative beliefs" (331). He argues that Belief Loss

seems just a truism about the way our psychology works, the kind of effect we routinely observe when someone becomes an atheist, or comes to believe that there is no such thing as phlogiston, or when you reveal to them that the elaborate anecdote you have been telling was a practical joke. (338)

But if nihilism is true, we would still care about all kinds of things: a scenario in which nothing really matters and yet we care about all kinds of things, such as suffering, is exactly the scenario that obtains right now if nihilism is in fact true. Is it really plausible that, in such a scenario, we would stop caring about the things we actually care about if we came to believe (correctly) that nihilism is true? If we came to believe that nothing matters in any sense whatsoever, then perhaps we really would stop caring about anything at all. Yet it seems quite possible to hold, like

Street, that “[n]othing matters, ultimately, independently of the attitudes of beings who take things to matter” (that nothing has “final value,” as Kahane puts it), but that nonetheless, “plenty of things *matter*” (2016b, 121).<sup>38</sup> And I doubt that holding such a position, correctly or incorrectly as the case may be, prevents one from caring about such things as justice and human suffering.<sup>39</sup>

Kahane insists that our subjective concerns co-vary closely with our evaluative beliefs, and consequently that “once we conclude that all of our evaluative beliefs are false, we should also largely lose the corresponding subjective concerns and motivations” (339-340). That may well be true, but it is far less clear that coming to believe in nihilism in Kahane’s specific sense—the sense according to which nothing has final value—will lead us to reject all our substantive evaluative beliefs as false: “to no longer hold that suffering is bad, virtue good, freedom desirable, etc” (340). We might be able to believe that suffering is bad without believing that it is ultimately bad, or that it is bad completely independently of us

---

38. Or even, as Street says, that “plenty of things ‘really’ matter, if we allow, as I think we should, that existing independently of a subject’s point of view of the world is not the only way of being ‘real’” (121). What Kahane seems to mean when he talks about things mattering, however, is precisely the “robustly attitude-independent sense” of “(really) matters” that Street rejects.

39. As a Robust Metanormative Naturalist, I of course do believe that some things matter in the robustly attitude-independent sense. Yet because of the rather abstruse metaphysics involved in the question of whether anything matters in this sense (see my own discussion of reductionism in Chapter 8, for instance), I don’t think that believing (incorrectly) that nothing matters in this sense prevents one from caring about things which really do matter (in this sense).

and our attitudes, or the like.<sup>40</sup> We will then be able to continue caring about suffering even if we come to believe that nihilism (in Kahane's sense) is true. Unlike Kahane's Wager, Parfit's Wager is a purely conceptual argument, and it is thus not answerable to the contingencies of our normative or evaluative psychology. I therefore believe this argument can succeed where Kahane's Wager fails.

### 3.2. Normative deliberation, IBE and bootstrapping

If we can really use normative deliberation to justify normative deliberation in the way that I am suggesting, normative deliberation must be able to pull itself up by its own bootstraps from an epistemic point of view. This need not be a problem: crucial to the Deliberative Indispensability Argument is an analogy between normative deliberation and the form of reasoning known as inference to the best explanation (IBE), and it would appear that IBE has just as much of a need to pull itself up by its own bootstraps as normative deliberation (at least in so far as IBE and normative deliberation are to count as reliable or epistemically justified to any extent).<sup>41</sup>

---

40. We saw in the previous chapter (Section 2.4) that there are no normative truths without *mind-independent* normative truths. But evaluative language clearly admits of an expressivist usage that doesn't commit us to the existence of (mind-independent) normative truths. We will still be able to use utterances like, "suffering is bad" and "suffering matters" to express our subjective concern with ameliorating suffering if nihilism in fact turns out to be true.

41. One might regard this analogy between normative deliberation and IBE as a version of the so-called "companions in guilt (or innocence)" strategy for defending metanormative realism: both the justificatory and the explanatory forms

It is instructive how Olson (2014), as a normative sceptic, actually responds to Enoch. He rejects the Deliberative Indispensability Argument, not for the same reasons that we have rejected it, but on the grounds that the only indispensability arguments that work

are precisely those that are truth-tracking, i.e. those that tend to render us having true beliefs. Employing IBE has rendered us having true beliefs about such things as protons and planets, and about mundane matters such as whether it has rained recently. This is because the world is explanation-friendly. (176)

By contrast, the world is not (normative) deliberation friendly, according to Olson, because there are no (mind-independent) normative truths. This response naturally takes IBE, but not normative deliberation, to be truth-tracking or epistemically justified (or reliable). But what if Enoch were to ask Olson how he knows that IBE is truth-tracking or epistemically justified? Judging by the above passage, he would presumably answer that IBE has led to us having true beliefs in the past, or that the world seems to be explanation-friendly. But to get from these observations to the conclusion that IBE is truth-tracking or epistemically justified, he will have to employ IBE, in which case he is leveraging IBE in its own defence. Olson suggests that the explanation-friendliness of the world is the *best explanation* for the success of IBE, but of course IBE can only be truth-tracking if the world is indeed explanation-friendly; if we were not justified in (or had no evidence for) taking the world to be explanation friendly in the first place, we would not be justified in taking IBE to be

---

of rational inquiry arguably suffer from the same Original Sin of ultimately having to pull themselves up by their own bootstraps (epistemically speaking).

truth-tracking. And we need to rely on IBE to support or license the conclusion that the world is explanation-friendly. So in terms of the justification, rationale or evidence for IBE, there appears to be an intractable element of epistemic circularity. This does not mean that we are not justified in accepting IBE; Enoch's point, however, is that normative deliberation is no less problematic in terms of its ultimate rationale or justification than IBE, and thus that we are no less justified in taking the deliberative indispensability criterion to be a sound basis for commitment to certain kinds of truths than the explanatory indispensability criterion.<sup>42</sup> Since explanatory indispensability is fundamentally no better off than deliberative indispensability, the analogy between the two stands, and IBE cannot be used to refute the commitments of normative deliberation.

The exact same thing can be said about a principle of parsimony like Occam's Razor. As Yudkowsky (2007b) explains:

If two hypotheses fit the same observations equally well, why believe the simpler one is more likely to be true? You could argue that Occam's Razor has worked in the past, and is therefore likely to continue to work in the future. But this, itself, appeals to a prediction from Occam's Razor ... Indeed,

---

42. Under some models of knowledge, the kind of bootstrapping I am claiming is involved in IBE would not come out as justified, and so the kind of bootstrapping involved in normative deliberation would not come out as justified either. But if coherentism is the correct theory of epistemic justification, this kind of bootstrapping sometimes will be justified, and IBE and normative deliberation are both prime candidates for justified cases of bootstrapping. If, on other hand, some form of foundationalism is true, then there will be some foundational, self-evident truths, and if there are some self-evident non-normative truths, there will probably be some self-evident normative truths as well. In any case, there is good reason to believe that normative deliberation is just as basic as IBE, whether IBE can bootstrap itself or is simply self-evident in terms of its justification.

it seems that there is no way to *justify* Occam's Razor except by *appealing* to Occam's Razor.

A scenario in which Occam's Razor will continue to work in the future because it has worked in the past is indeed a simpler one than a scenario in which Occam's Razor has worked in the past and then suddenly, for some reason or other, stops being as effective in the future,<sup>43</sup> and thus, *by Occam's Razor*, the former scenario is more likely to obtain than the latter. Furthermore, Occam's Razor has worked in the past (presumably) because it has led to us having true beliefs in the past: using Occam's Razor has often led us to select those hypotheses which actually are (most likely to be) correct. But how do we know that the hypotheses we have selected on the basis of Occam's Razor are actually the correct ones (or the ones that are most likely to be correct)? It would seem that there is no way to confirm the truth of these hypotheses in a way that is independent of the original appeal to Occam's Razor.

It is reasonably clear that the sceptic who, like Olson, relies on IBE and Occam's Razor to argue against metanormative realism needs these things to pull themselves up by the bootstraps from an epistemic point of view. It is therefore open to the realist to argue that normative deliberation (and indeed the entire normative domain) can bootstrap itself in an equally legitimate way. The normative sceptic, however, can give an informative account of how this bootstrapping is supposed to work in the case of IBE: he can tell us how the way the world is explains how certain

---

43. Yudkowsky points out that the hypothesis: "'Occam's Razor works up to October 8th, 2027 and then stops working thereafter' is more complex, but it fits the observed evidence equally well."

commitments of ours are likely to track the truth, and by using IBE to support IBE, he can tell us how and why IBE actually works. Normative deliberation and a justified belief in (mind-independent) normative truths probably need to pull themselves up by their own bootstraps as well. IBE and normative deliberation both appear to be what Enoch calls “basic belief-forming methods” (2013, 58), i.e. methods of forming justified beliefs which nonetheless do not admit of further epistemic justification by virtue of anything more fundamental than themselves. So unless the realist can give an informative account of the kind of bootstrapping involved in normative deliberation, he cannot claim that normative deliberation is epistemically on par with IBE. The realist should be able to show that deliberation can pull itself up by the bootstraps in a way that is epistemically just as respectable as IBE.

This is where our second argument for metanormative realism—Parfit’s Wager—comes in. For Blaise Pascal, we ought rationally to believe in the Christian God, although His existence is plainly uncertain, because wagering that God exists is the best bet in *prudential* terms (Hájek 2018). In a similar vein, Parfit’s Wager purports to show that we ought rationally to believe in the existence of normative reasons and truths, because wagering as much is the best bet in *normative* terms. But fortunately for us, Parfit’s Wager does not suffer from the same defects as Pascal’s original Wager. It is not vulnerable to a “What about the other Gods?” objection, showing that the prudential value of believing in the Christian God is cancelled out when we consider, say, the possibility of a God that doesn’t want us to believe in

Him, and will vindictively punish us if we dare to do so. Nor does it require us to actively *try* to believe something that we might just find it too difficult to believe. Instead, it simply contends that we will come to understand, as a conceptual matter, that we have a categorical reason to believe in normative reasons (and a normative reason to believe in categorical reasons) when we reflect properly on the *possibility* that there could be such reasons and what this possibility means for us and for what (if anything) we ought rationally to believe.<sup>44</sup>

At first glance, this argument might strike the reader as viciously circular, in spite of the analogy with IBE and its need to pull itself up by the bootstraps (epistemically speaking). How can we wager for the existence of normative reasons because it is the best bet in *normative* terms unless we already know that normative reasons actually exist? The answer is that when we make the Wager, we are comparing epistemically possible worlds in terms of the normative reasons (or lack thereof) in those worlds, without assuming, at that stage, that the actual world contains any normative reasons itself. For all we know at this stage, we might be in the neutral world (in normative terms), by which I mean a world in which there are no normative reasons and in which, therefore, everything is normatively neutral, in the sense that nothing has any normative property or status (positive or negative). What we are doing (or ought to be doing) is giving ourselves the best chance of

---

44. The categorical reason we have to believe in normative reasons is an epistemic reason (a reason for a belief), but the normative reasons we have categorical (epistemic) reason to believe in might be other kinds of reasons (such as practical reasons), and they might be non-categorical as well as categorical reasons.

being in a good world in normative terms, that is, a world in which normative reasons do exist and in which we respond appropriately to those normative reasons (in that world). In so doing, we are *not* already assuming, in a question-begging manner, that we are not in fact in the neutral world. The existence of normative reasons (in the actual world) is a conclusion we can then go on to draw on the basis of the Wager.

### 3.3. *The argument*

Without further ado, therefore, here is the passage from Parfit (2011b) containing what I refer to as Parfit's Wager:

If we believe that there are some irreducibly normative truths, we might be believing what we ought to believe. If there are such truths, one of these truths would be that we ought to believe that there are such truths. If instead we believe that there are no such truths, we could not be believing what we ought to believe. If there were no such truths, there would be nothing that we ought to believe. Since

(D) it might be true that we ought to believe that there are some irreducibly normative truths,

and

(E) it could not be true that we ought not to have this belief,

we can conclude that

(F) we have unopposed reasons or apparent reasons to believe that there are such truths,

so that

(G) this is what, without claiming certainty, we ought rationally to believe.<sup>45</sup> (619)

---

45. Parfit also makes the claim that "we could not have reasons to believe that there are no irreducibly normative truths, since the fact that we had such reasons would itself have to be one such truth," and the claim that, "if we ought to have some belief about this question [whether or not there are irreducibly normative truths], this [that there are such truths] is what we ought to believe" (618-619). I reject the first of these two claims. As Streumer (2017) points out, "there can be

And if we ought rationally to believe (even if we have no right to be certain) that there are some irreducibly normative truths, I think we can conclude that normative reasons and truths (probably) do exist: the unopposed or apparent reasons we have to believe in irreducibly normative truths would themselves be some such reasons, and the truth that we have such reasons would itself be one such truth.

I don't think the above passage establishes the existence of *irreducibly* normative reasons and truths, however. If we speak just in terms of normative reasons and normative truths, and leave it open whether or not these reasons and truths are metaphysically reducible, presumably to naturalistic or empirical truths, we will get the conclusion that we ought rationally to believe (without claiming certainty) that there are some normative reasons and normative truths. Of course, it may be quite impossible to reduce normative reasons and normative truths to anything more fundamental than themselves, and Parfit certainly has other arguments for the thesis that such reasons and truths must be irreducible.<sup>46</sup> But as

---

reasons to believe a theory [such as the normative error theory] that is in fact false but we do not know is false" (155). If we do not know that there are any (irreducibly) normative truths, there can be reasons to believe that there are no such truths, even if this would mean that there actually are such truths. Yet I think the second claim is clearly true: if there are no normative truths, it could not be the case that we ought to believe that there are no such truths (because this would itself have to be one such truth), but if there are such truths, it certainly could be the case that we ought to believe that there are such truths.

46. Such as the "Triviality Objection" against normative naturalism, to which he devotes a whole chapter (2011b, 328-356). We will discuss the Triviality Objection in Chapter 6 (Section 6.4). And I will defend non-reductive normative naturalism over the course of Chapters 7 and 8.

with the Deliberative Indispensability Argument in the previous chapter, we will set this issue aside for now and leave it completely open whether normative reasons and truths are irreducible or not.

### 3.4. *The Normative Matrix and Normative Superdominance*

We can present Parfit's Wager in the form of a decision matrix (à la Pascal's Wager):

	<b>There are some normative reasons and truths</b>	<b>There are no normative reasons and truths</b>
<b>Believe that there are some normative reasons and truths</b>	Believe what we ought to believe	Normatively neutral
<b>Believe that there are no normative reasons and truths</b>	Believe what we ought not to believe	Normatively neutral
<b>Suspend judgement</b>	Fail to believe what we ought to believe	Normatively neutral

According to this matrix (let's call it "the Normative Matrix"), if there are no normative reasons and truths, there is nothing that we ought to believe (or ought not to believe), and therefore, whatever we believe (or don't believe) in such a scenario is normatively neutral: it neither has any normative property or status (positive or negative) itself, or involves the possession of any normative property or status by anything whatsoever. In a scenario such as this, it doesn't matter normatively what we believe (or don't believe). If, on the other hand, normative reasons and truths

actually do exist, then, if we also believe that there are some such reasons and truths, we will be believing what we ought to believe. Yet if, in this same scenario, we instead believe that there are no such reasons and truths, we will be believing what we ought *not* to believe: if we ought to believe that there are some such reasons and truths, it follows that we ought *not* to believe that there are *no* such reasons and truths. And what if we simply suspend judgment altogether, believing neither that there are nor that there aren't any such reasons and truths? If there are no normative reasons and truths, it won't matter normatively whether or not we suspend judgment on this question. But if there are some such reasons and truths, then although we might avoid believing what we ought not to believe by suspending judgment, we will also be *failing* to believe what we ought to believe, because, in such a scenario, we ought to believe that there are some such reasons and truths.

One thing this matrix assumes is that, in the scenario in which normative reasons and truths exist, it is also the case that we ought to believe that there are such reasons and truths (as Parfit says, “[i]f there are such truths, one of these truths would be that we ought to believe that there are such truths” (619)). But are we really entitled to assume this? Couldn't it instead be the case that, even though there are such things as normative reasons and truths, we have no reason at all to believe in them?<sup>47</sup> Suppose that there are normative reasons and truths. In this scenario, there will be some things that we have normative reasons to believe. And if we ever

---

47. After all, some things can obviously exist without us necessarily having any reason to believe in them.

*have reasons to believe* that we have reasons to believe any of the things that we in fact have reasons to believe, we will also (thereby) have reasons to believe that there are normative reasons and truths; we will then be able to infer the existence of normative reasons and truths straightforwardly from the fact that we have these reasons. And if (in this same scenario) we ever had any countervailing reasons to believe that there are no such reasons and truths, these very reasons that we have would themselves indicate that normative reasons and truths actually exist.

If there are any normative reasons and truths, therefore, it follows that we ought to believe that there are such reasons and truths—unless, that is, we never have any reasons whatsoever to believe that we have any reasons to believe any of things that we nonetheless do have reasons to believe. And while this is a logically possible combination, it is not exactly a plausible one.<sup>48</sup> If there are some normative reasons and truths, there are surely some things that we have normative reasons to believe, and although there will no doubt be cases in which we have reasons to believe something without knowing that we have these reasons, in all probability, there will also be cases in which we *do* know that we have reasons to believe what

---

48. Especially when we consider that, just because there are some normative reasons and truths does not mean that non-naturalism has to be true about them, such that there might have to be a particular epistemological problem about our access to a metaphysically *sui generis* domain of normative truths that doesn't arise for other areas of knowledge. I will defend a naturalistic account of normative knowledge in Chapter 7.

we actually have reasons to believe.<sup>49</sup> If I have reasons to believe that I have a hand, I probably also *have reasons to believe* that I have reasons to believe that I have a hand (because I also have reasons to believe that I have eyes that work correctly, for instance). And if we ever know that we have normative reasons to believe something, we are thereby in a position to know that there are some normative reasons, simply by inference from the fact that we have these reasons. It does seem to be the case, therefore, that if there are some normative reasons and truths, this is what we ought to believe.<sup>50</sup> This means that, in accordance with the Normative Matrix, if we ought to believe that there are some normative reasons and truths, and we instead believe that there are no such reasons and truths, we will be believing what we ought not to believe; and if we suspend judgement on the matter, we might avoid believing what we ought not to believe, but we will still be failing to believe what we ought to believe.

So taking all these possible outcomes into account, what can we say follows from the Normative Matrix? Does it really show that we ought rationally to believe

---

49. The complete normative sceptic, of course, will not accept that there are such cases. But such a sceptic would probably not also be a metanormative realist, and it hardly seems tenable to be both a sceptic and a realist about normativity at the same time.

50. We should qualify this by saying that we ought not to hold this belief with certainty, to allow for the *possibility* that, even though normative reasons exist and we sometimes have normative reasons for belief, there are absolutely no cases in which we know that we have such reasons. According to claim Parfit's (G), "this is what, *without claiming certainty*, we ought rationally to believe" [my emphasis] (619). The conclusion that Parfit's Wager delivers is precisely that there are normative reasons, and that we ought to believe as much—but without certainty.

that there are some normative reasons and truths? It may establish a version of Parfit's claim (F): that there are *apparent* reasons to believe in normative reasons and truths, but this will not necessarily entail or support the conclusion that we ought rationally to believe in such reasons and truths (even if we avoid claiming certainty).<sup>51</sup> For this to follow, we need there to be *actual* reasons to believe that there are some such reasons and truths. And I don't think the Normative Matrix on its own is enough to entail that we ought rationally to believe that there really are some normative reasons and truths.

To get to this conclusion from the Normative Matrix, we need one more premise that Parfit himself does not supply. This premise is a principle I call "Normative Superdominance." The Normative Superdominance principle states that:

If a particular decision or response (such as a forming a certain belief) *normatively superdominates* all its available alternatives, then we have *at least some* normative reason to make that decision or to respond in that way (for instance, by forming the relevant belief).

---

51. Talk of rationality as I am engaging in it here (and as Parfit does in the original passage) is interchangeable with talk of reasons. This is largely a matter of stipulation: there is a sense of "rationality" which does not inherently involve normative reasons, meaning instead something like "reason," which itself refers to the mental abilities that lead us to have true beliefs. Even if there are no normative reasons and truths, these mental abilities of ours would still exist, and thus reason or rationality, in that sense, would still exist in the absence of normative reasons. But this is obviously not the relevant sense to use in this context. As Parfit says, "the more important questions are not about rationality, but about reasons" (454). When we are asking whether or not we ought rationally to believe in normative reasons and truths, therefore, we are asking whether or not we have (normative) reasons to believe in such reasons and truths.

A decision or response “normatively superdominates” all its available alternatives if and only if

- 1) the *normatively best* possible outcome associated with this decision or response is *normatively better* than any of the possible outcomes associated with its alternatives,
- 2) the *normatively worst* (or least normatively good) possible outcome associated with this decision or response is *normatively equal to or better* than the *normatively best* possible outcome associated with any of its alternatives.

and

- 3) the *normatively worst* possible outcome associated with each one of its alternatives is *normatively worse* than the *normatively worst* possible outcome associated with this decision or response.

In other words, the particular decision or response in question not only has to provide us with the only chance we have of getting the normatively best possible outcome; there also has to be no way that we could get a normatively better outcome (better, that is, than we would *actually* get if we responded in that way) by responding in some alternative way instead (because the possible outcomes associated with any of the alternatives are none of them normatively better than *any* of the possible outcomes associated with this decision or response). And not only that, but this decision or response has to provide us with the only guarantee that we will avoid the normatively worst possible outcome: all of the alternatives must involve some chance of getting an outcome which is normatively worse than any of the outcomes we would get if we made the decision or response in question. If all three of these conditions obtain, it seems to be a conceptual truth that we have at least *some* normative reason to make that decision or to respond in that way, and

thus that it is not the case that we have absolutely no normative reason whatsoever to respond in that way.<sup>52</sup>

When I speak of a possible outcome being normatively better or worse than another possible outcome, I am defining a “normatively good” outcome as one in which we respond appropriately to normative reasons (for instance, in which we believe what we have normative reasons to believe, or do what we have normative reasons to do) and a “normatively bad” outcome as one in which we fail to respond appropriately to normative reasons (in which we fail to believe what we have normative reasons to believe, or in which we fail to do what we have normative reasons to do). I also define a “normatively neutral” outcome as one which is neither normatively good nor normatively bad (in these senses). An outcome where we neither respond nor fail to respond appropriately to normative reasons, because normative reasons do not exist and thus there are no normative reasons for us to respond or fail to respond to, is in this sense a normatively neutral outcome. Given these definitions, we can say that a normatively good outcome is, of course, normatively better than a normatively bad outcome. I also think we can say that a normatively good outcome is normatively better than a normatively neutral one: if

---

52. In support of this idea, we can observe that it is clearly a conceptual truth that, if we ought to make a certain decision or respond in a certain way, then we have decisive reason to make that decision or to respond in that way. This seems to suggest that, if we are more likely to be responding as we ought to respond, and less likely to be responding as we ought not to respond, by responding in a certain way (or by making a certain decision), compared to any other decision or response (or lack thereof) we could possibly make, and we know this, then we have, perhaps not *decisive* reason, but *at least some* reason to make that decision or to respond in that way.

one outcome is good (in some way) and another outcome is *not* good (in that same way), then, even if the latter outcome is not bad (in that way), it is reasonable to infer that the former outcome is better (in that way) than the latter outcome. And for the same reason, a normatively bad outcome is also normatively worse than a normatively neutral outcome (even though the normatively neutral outcome isn't normatively good).

With these conceptual resources in hand, we can clearly see by looking at the Normative Matrix that believing in normative reasons and truths normatively superdominates all its available alternatives. The normatively worst (or least normatively good) possible outcome involved in having this belief (the normatively neutral outcome) is as normatively good as the normatively best possible outcome involved in either having the contrary belief or in suspending judgement (the normatively neutral outcome in both cases). Believing in normative reasons and truths is the only response that has the normatively best possible outcome (believing what we ought to believe) associated with it; and the normatively worst possible outcome associated with either of the other two responses (believing what we ought not to believe or failing to believe what we ought to believe) is worse than the worst possible outcome associated with believing in such reasons and truths (mere normative neutrality). Since believing in such reasons and truths normatively superdominates both of its only possible alternatives, by the Normative Superdominance principle, we have at least some normative reason to believe in such reasons and truths. And if we have even *some* normative reason to believe in

such reasons and truths (no matter how tiny or weak this reason may be), it follows that such reasons and truths must therefore exist: the reason that we have to believe in such reasons and truths would itself be one such reason, and the truth that we have this reason would itself be one such truth. And if *this* is the case, we can indeed conclude that we ought rationally to believe that there are some normative reasons and truths.

That being said, the Normative Matrix does have as its precondition the potentially controversial assumption that normative reasons and truths are *possible*. If such reasons and truths are not in fact possible, it would make no sense to ask whether or not to believe in such reasons and truths, or whether or not one had *reason* to have this belief, and there would be no normatively good or bad possible outcomes involved either in having or not having this belief, so there would be no way of generating the Matrix in the first place. This point can also be expressed in terms of probability: if you assign probability 0 to the existence of normative truths and normative reasons—as a complete normative sceptic might—then Parfit's Wager will not give you a reason to change your mind and become a metanormative realist. One of the premises of the argument is claim (D): that it *might* be true that we ought to believe that there are some (irreducibly) normative truths (619). And the sceptic could well reject this premise, arguing that because there are no normative reasons and truths, we can effectively rule out from the start the possibility that we might be believing what we ought to believe by believing in such reasons and truths.

At this stage, the realist should not simply point out that normative reasons and truths are *logically* possible: the sceptic does not have to maintain that the negation of any normative proposition either is or reduces to a truth of logic. What the realist needs is for it to be *epistemically* possible that some normative propositions are true—for it to be the case that, when coming to Parfit's Wager, we can reasonably be uncertain or undecided about whether or not there really are any specifically normative reasons and whether or not we have any reason to believe in them. As far as the sceptic is concerned, the upshot of this is that, if he already has a strong argument against the existence of (mind-independent) normative truths, Parfit's Wager will not provide an answer to this argument, because this argument will be enough to establish that such truths are epistemically impossible. If some version of the Argument from Queerness, for instance, is sound as an argument against all normativity, then we would be able to *know* on the basis of this argument that there are no (mind-independent) normative truths, and Parfit's Wager would do nothing to change this state of affairs.<sup>53</sup> As far as the realist is concerned, however, the upshot is that the onus is very much on the sceptic to provide a knock-down argument against realism, on the basis of which we can know that there are no (mind-independent) normative truths, because if there is no such argument, then (mind-independent) normative truths will be epistemically possible, and claim (D) will

---

53. Olson, who accepts a version of the Argument from Queerness, would maintain that we do (or at least can) know that there are no (mind-independent) normative truths, because such truths would imply the existence of irreducibly normative favouring relations, and irreducibly normative favouring relations are queer. I will address this argument in Chapter 5 (Section 5.4).

therefore be true, thus generating the Normative Matrix and with it (via the eminently plausible Normative Superdominance principle) the conclusion that there really are some normative reasons and truths.<sup>54</sup> Even if Parfit's Wager does not itself qualify as a knock-down argument against normative scepticism, it does succeed in very much shifting the burden of proof onto the sceptic, who must show that he *knows* (with something approaching certainty) that there is no such thing as (mind-independent) normative truth. It is still a significant result that, if normativity is even possible, it (probably) does actually exist.

In our context, remember, we are suggesting that normative deliberation can pull itself up by its own bootstraps in a way that is epistemically on par with the way that inference to the best explanation can pull itself up by its own bootstraps. Of course, IBE must pull itself up by the bootstraps in an explanatory way: a way that pivots on the notion of an explanation.<sup>55</sup> Normative deliberation, on the other hand, must pull itself up by the bootstraps not in an explanatory way, but in a normative way: a way that pivots on the notion of what we ought rationally or have (normative) reason to believe. Using Parfit's Wager, we have deliberated our way to the conclusion that we ought rationally to believe in the existence of normative

---

54. Plus the conclusion that (some of) these normative reasons and truths must be mind-independent, taking into account the truth of premise (1) of the Deliberative Indispensability Argument (which we established in the previous chapter).

55. Yudkowsky offers such an explanation for why Occam's Razor works in practice when he points out that "we live in a simple world, a low-entropy universe in which there are short explanations to be found" (2007b). This is surely the same basic reason why IBE works in practice as well.

reasons and truths. This suggests that normative deliberation can pull itself up by the bootstraps in an epistemically respectable way. Of course, this argument does not prove the existence of normative reasons and truths by means of IBE. But then it would appear that the felicity of IBE cannot be proven by means of normative deliberation. All-things-considered, therefore, deliberation and IBE do seem to be on par from an epistemic point of view, taking each basic belief-forming method on its own terms (as it makes most sense to do). The normative sceptic who attempts to refute the commitments of deliberation by appealing to IBE would be akin to a sceptic of IBE who attempts to refute this basic belief-forming method by appealing to normative deliberation. In both cases, this is the wrong approach. Both basic belief-forming methods have some need to bootstrap themselves and both seem equally capable of doing so. As a result, the commitments of normative deliberation are just as respectable as the commitments of IBE.

### *3.5. From normative reasons to categorical reasons*

In Chapter 1 (Section 1.1), I stated that normative reasons depend on categorical reasons. I will now explain why the existence of normative reasons implies the existence of categorical reasons. And if I am correct that Parfit's Wager establishes the existence of normative reasons, this means that it also establishes the existence of categorical reasons. Categorical reasons are the kind of normative reasons which apply to agents independently of their involvement in certain rule-governed or goal-orientated activities. As Olson (2011) elaborates:

To say that there are categorical reasons for some agent, A, to behave in some way,  $\varphi$ , is to say that there is reason for A to  $\varphi$  irrespective of whether A's  $\varphi$ -ing would promote satisfaction or realization of some of A's desires or aims, or promote the fulfilment of some role A occupies, or comply with the rules of some activity A is engaged in. (64)

Such reasons need not be just reasons to behave in some way; we might have categorical reasons to respond in other ways, such as by forming certain beliefs, so long as the reasons we have to respond in those ways apply to us irrespective of psychological or social phenomena such as our desires, ends, or roles.<sup>56</sup>

Some theorists believe that there are normative reasons but no categorical reasons: that all our normative reasons, whether for action or belief, are dependent in some way on our aims or desires. I don't think this *Internalism about reasons*,<sup>57</sup> as it is often called, represents a coherent position. As Scanlon (2014) points out, "the claim that a person has [normative] reason to do what will promote the satisfaction of his or her desires is itself a normative claim. Indeed, it is an 'objective' normative claim, since it does not itself depend on what people desire, or on what aims they have" (16-17). As we saw in the previous chapter (Section 2.3), Olson argues that we

---

56. Categorical reasons are also sometimes called *external* reasons. Street defines external reasons as "reasons for agents to do things even in cases where there is no sound deliberative route from their own subjective motivational set to the conclusion that it is a reason" (2016b, 132). I think this is merely a terminological difference: categorical or external reasons are those reasons that apply to us independently of our own or anyone else's motivations.

57. Internalists about reasons are those who "agree that there are reasons in the standard normative sense, but deny that there are any normative reasons of a certain sort—namely ones that exist independently of whether there is a sound deliberative route to the conclusion that they exist from the subjective motivational set of the agent whose normative reasons are in question" (131).

only have hypothetical reasons for action and belief, and that these hypothetical reasons are in no way normative: they are not “properly understood in terms of the counting-in-favour-of relation,” as he puts it.<sup>58</sup> In other words, we may have reasons to respond in certain ways, in the purely psychological sense that responding in these ways is likely to satisfy our desires, but facts about our desires and the likely means to their satisfaction do not actually call for or count in favour of any responses on our part. Claims that say that certain features of the world call for or count in favour of certain responses do not “reduce to empirical claims about agents’ desires and (actual or believed) efficient means of bringing about the satisfaction of these desires,” and are therefore, in Olson’s view, systematically false. For instance, it might be true to say “that there was reason for Hitler to invade Britain during World War II,” but if so, this “means nothing more than that Hitler wanted to win the war and had he invaded Britain he would have been more likely to do so” (2011, 78). It does not mean that Hitler’s invading Britain was called for or counted in favour of in any respect whatsoever; and if it did mean this, the claim that Hitler had this reason would be false, just like, for Olson, any normative claim implying the existence of categorical reasons.

This suggests that a sceptic about categorical reasons should also be a sceptic about all normative reasons. Olson tells us that the sceptic about categorical reasons

---

58. That is, the concept of a hypothetical reason is not related to or acquired from the conscious experience of certain features of the world calling for or counting in favour of certain responses on our part, but simply from the observation that agents act in order to achieve certain ends.

“cannot hold that there is a transcendent norm to the effect that agents take (what they believe to be) the means to their ends, for that would mean that the error theorist is committed to their being categorical reasons after all” (77).<sup>59</sup> If Hitler did have a normative (hypothetical) reason to invade Britain, this implies that he would have had a categorical reason to take the means to his ends: if he didn’t have a normative reason to take the means to his ends, a reason which did not *itself* depend on the ends or desires he happened to have, he wouldn’t have had a normative reason to do the specific thing (invading Britain) that would have been conducive to the satisfaction of his actual desire (winning the war). And since Olson believes that there are no categorical reasons, he concludes that Hitler had no normative reason to invade Britain—and that no one ever has any normative reasons at all, for that matter.

I think this example shows that, if there are any normative reasons at all, there must be at least one transcendent norm in virtue of which we have these reasons, and which applies to us completely independently of our aims or desires. Even if there is only one transcendent norm, according to which we ought to take the most efficient means towards our ends, we would still have a categorical reason to comply with this norm and thus to take the most efficient means towards our ends. The Internalist might attempt to escape this problem by making an appeal to meta-

---

59. Olson makes this observation in the context of defending moral error theory: the view that moral claims are or entail claims about categorical reasons, but there are no categorical reasons in reality and consequently all moral claims are false (62). If Parfit’s Wager really does establish the existence of categorical reasons, it significantly undermines the case for moral error theory in the process.

desires. If we have a meta-desire to satisfy our first-order desires as thoroughly as possible, for instance, this meta-desire might give us a reason to take the means to our ends that is normative and yet hypothetical, rather than categorical. But I think Olson could easily respond to this by saying that this simply pushes the problem back one step: if a meta-desire to take the most efficient means towards one's ends gives one a *normative* reason to do just that, it still follows that there must be a transcendent norm to the effect that agents take the most efficient means to towards the fulfilment of their (meta-)desires, a norm which inescapably implies the existence of categorical reasons.

But even if the Internalist can somehow accommodate normative reasons without categorical reasons (by appealing to meta-desires or otherwise), I think that Parfit's Wager still establishes the existence of categorical reasons in its own right. By Parfit's Wager, the reason we have to believe in normative reasons and truths does not come from any relation this belief stands in to our desires, but from the conceptual truth that believing in such reasons and truths normatively superdominates not believing in such reasons and truths. In other words, it follows from the epistemic possibility of normative reasons plus the Normative Superdominance principle that we not only have a normative reason to believe in normative reasons and truths, but that this normative reason is also a categorical reason. The only way of escaping this conclusion, I believe, is by denying one (or both) of these premises. And I don't think someone who denies either of these premises could consistently hold that there are any normative reasons at all.

### 3.6. *The Deliberative Indispensability Argument revisited*

We are now in a position to see how the Deliberative Indispensability Argument, supplemented by the kind of Pascalian reasoning involved in Parfit's Wager, establishes the existence of mind-independent normative truths. The correct formulation of this argument states that:

- (1\*) If there were no mind-independent normative truths, we would not have sufficient *normative* reason to engage in *normative* deliberation.
- (2\*) We *do* have sufficient normative reason to engage in normative deliberation.
- (3\*) Therefore, there are mind-independent normative truths.

Premise (2\*) is of course the crucial premise of this argument. And to illustrate further why this premise is true, we can use the following matrix:

	<b>There are some normative reasons and truths</b>	<b>There are no normative reasons and truths</b>
<b>Sometimes engage in normative deliberation</b>	Normatively good	Normatively neutral
<b>Never engage in normative deliberation</b>	Normatively bad	Normatively neutral

In the scenario in which there are no normative reasons and truths, there is of course no normative difference between engaging and not engaging in normative deliberation: there are no normative truths for us to discover, so it does not matter normatively whether or not we try to discover such truths. But suppose there really are some normative reasons and truths. If there are some normative reasons and

truths, a scenario in which we respond appropriately to those normative reasons and truths (for instance, by doing what we ought to do) will in normative terms be better, all else being equal, than a scenario in which we do not respond appropriately to those normative reasons and truths (such as by not doing what we ought to do). And if, in such a scenario, we engage in some degree of normative deliberation, at least some of the time, we are surely more likely to discover what the normative truths are, and thus more likely to respond in a normatively appropriate manner to those truths, than if we never engage in any normative deliberation whatsoever. We can therefore say that, if there are some normative reasons and truths, one of these truths would be that we ought to engage in (some degree of) normative deliberation.

As such, if there are some normative reasons and truths, and we do sometimes engage in normative deliberation, we are doing what we ought to do in that particular scenario, and are thus in the normatively good case, all else being equal. If we never engage in normative deliberation in that same scenario, we are failing to do what we ought to do, and are thus in the normatively bad case (all else being equal). And if there are no normative reasons and truths, it doesn't matter normatively whether or not we ever engage in normative deliberation. Therefore, when it comes to deciding whether or not to engage in normative deliberation to any extent whatsoever, we can say that, if we choose to engage in normative deliberation at least some of the time, we will find ourselves either in the normatively good case or the normatively neutral case. If we instead choose never to engage in normative

deliberation, we will get either the normatively neutral case or the normatively bad case. Normatively speaking, the combination of possibilities that includes the normatively good case and the normatively neutral case clearly trumps the combination of possibilities that includes the normatively neutral case and the normatively bad case. It certainly seems normatively better not to run the risk of getting the normatively bad case, especially when, by doing so, we will at best get the normatively neutral case, and to allow for the possibility of getting the normatively good case (when at worst, all we will get is the normatively neutral case). And thus it follows that, normatively speaking, the decision to engage in some degree of normative deliberation trumps the decision never to engage in any normative deliberation at all. To say the least, therefore, we do seem to have some normative reason to engage in normative deliberation.

So long as normative reasons and truths are epistemically possible, we can clearly see that the decision to engage in some degree of normative deliberation (at least some of the time) normatively superdominates the decision never to engage in any normative deliberation whatsoever. If the Normative Superdominance principle holds, it follows that we have at least some normative reason to engage in normative deliberation. We must, therefore, be in the scenario in which there are such things as normative reasons and truths. And from this it follows that we not only have some reason, but *sufficient* reason to engage in normative deliberation (because we ought to engage in normative deliberation if there really are some normative reasons and truths). This entails that premise (2\*) of the Deliberative Indispensability Argument

is true, and when combined with the truth of premise (1\*), gives us the conclusion that there are mind-independent normative truths.

### 3.7. Conclusion

If IBE and Occam's Razor can be relied upon at all as means of acquiring knowledge, the kind of epistemic bootstrapping involved in them must not be completely fallacious; and this is the very same kind of bootstrapping that (I have argued) is also involved in Parfit's Wager and the Deliberative Indispensability Argument. Another way to put this point is to say that behind (the use of) IBE and Occam's Razor is the assumption that the universe is a (relatively) simple place: a place in which simple explanations of phenomena are more likely to be correct than more complicated ones. This assumption may be highly plausible. But then so, I believe, is the Normative Superdominance principle I have appealed to in defence of the existence of normative reasons and truths: if the possible outcomes involved in a certain decision or response are so much normatively better on balance than those involved in every possible alternative, it seems equally reasonable to assume that we have *at least some* normative reason to make that decision or to respond in that way.

We have seen from our discussion of premise (1) of the Deliberative Indispensability Argument in the previous chapter that normative deliberation is committed to mind-independent normative truths, in the sense that it aims at discovering normatively non-arbitrary answers to normative questions, and that there would be no such answers to be had without mind-independent normative

truths. As such, although there may still be some (even sufficient) *non-normative* reason to engage in normative deliberation in the complete absence of such truths, there would not be any *normative* reason to do so. What Parfit's Wager shows is that there *is* sufficient normative reason to engage in normative deliberation. And if we have sufficient normative reason to engage in normative deliberation, and we can use normative deliberation to arrive at the conclusion that there are some things we ought rationally to believe, normative deliberation must sometimes be able to achieve its aim of discovering normatively non-arbitrary answers to normative questions. There must, therefore, be mind-independent normative truths, and normative scepticism must be false. Having finally got what we came for over the course of the last two chapters, in the next chapter, we will consider the ontological question and complete our case for Robust Metanormative Realism.

#### 4. Ontological Implications

There probably are at least some (mind-independent) normative truths: that we have normative reason to believe in such truths is (probably) one of them. And since, therefore, a realist metanormative theory is called for, we need to ask how these (mind-independent) normative truths, in McPherson's phrase (2011), "fit within our best general account of the world" (223). To do this, we first need to ask what kind of ontological implications these truths possess, and thus whether or not we should opt for a more or less robust version of realism.

As we observed in Chapter 1, some theorists, such as Parfit (2011b), think that normative truths "have no positive ontological implications" (479). Parfit explains that normative properties "need not exist either as natural properties in the spatio-temporal world, or in some non-spatio-temporal part of reality" (486). Others, such as Enoch (2013), think that they do have ontological implications, and that metanormative realists should not disavow the robust ontological commitments of their view (7). The minimalist view (that normative truths do not have ontological implications or commitments) seems more parsimonious than the fully robust view, but this may not be the case if the Robust Realist can give an adequate naturalistic account of the ontological implications or commitments that normative truths are supposed to possess. If normative naturalism is not in fact plausible, however, then

minimalism *will* have the advantage of being more parsimonious than Robust Realism. So there is something of a burden on the Robust Realist (who is not a naturalist) to explain what exactly is wrong with normative minimalism. The burden on the minimalist (assuming that the Robust Realist cannot meet his burden) is to refute normative naturalism and establish the autonomy of the normative domain; it is the essence of the minimalist view (or “quietist realism,” as McPherson calls it (2011, 224)) that “[n]ormative truths ... constitute a distinct realm and need no natural or special metaphysical reality in order to have the significance we commonly grant them” (Scanlon 2014, 52).<sup>60</sup>

In chapters 6 and 7, we will find out if normative naturalism can hold up to scrutiny. In this chapter, I will argue—against minimalism and for Robust Realism—that normative truths *do* have ontological implications, on the grounds that normative facts and properties seem to possess some kind of causal efficacy and standing: our beliefs and actions are often counterfactually *sensitive* to differences in the normative facts, and to differences in the instantiation of normative properties. Some normative facts may be fundamental and necessary, but most of the normative facts that actually obtain at any given time are contingent, and if these contingent

---

60. Judging by this remark, Scanlon seems to agree with Parfit that normative truths do not have ontological implications, whether natural or “metaphysical” (perhaps meaning supernatural or at least non-spatio-temporal), but he in fact denies that his view is “minimalist,” because it “aims to give normative ... statements exactly the content and ‘thickness’ that they require when taken literally: no more and certainly no less” (28). In our sense of “minimalism,” however, the minimalist view is not minimalist about the *significance* (or content) of normative truths, but only about the *ontology* of such truths. So I will continue to classify Scanlon as a minimalist.

normative facts were different, certain non-normative facts (such as facts about our beliefs and actions and their consequences) would be also different as a result. This indicates that normative facts and properties have an influence on the real world—they are not epiphenomenal—and thus that there are “worldly” normative facts, as the Robust Realist believes, in addition to the “propositional” normative facts in which the minimalist believes.<sup>61</sup>

I will begin, in Section 4.1, by providing some motivation for the minimalist view that fundamental normative truths do not need “truthmakers” because they are exclusively *a priori*. In Section 4.2, I will draw upon an analogy with modal and mathematical truths, combined with D. M. Armstrong’s account of the truthmakers for these truths, to argue that, even if fundamental normative truths are *a priori*, this in itself would not exempt them from a general need for truthmakers. Then, in Section 4.3, I will endorse Parfit’s (and Skorupski’s) observation that there are two senses of the term “exist,” an ontological and a non-ontological sense, arguing that what demarcates the two senses is that the former implies some kind of causal standing whereas the latter implies the complete absence of causal standing, and that the central issue between the minimalist and the Robust Realist is therefore whether or not normative properties have some kind of causal standing. In Section 4.4, I will consider and reject Enoch’s argument that minimalism is incompatible with the existence of mind-independent normative truths because it is just a

---

61. See Chapter 1 (Section 1.2) for the distinction between the “worldly” and “propositional” senses of “fact.”

notational variant of *fictionalism*. And in Section 4.5, I will make the argument that normative properties *do* have some kind of causal standing by virtue of our sensitivity to them. Finally, in Section 4.6, I will reject the potential counterargument that normative properties do not actually have any causal standing because they supervene on natural properties, and thus it is these natural properties which actually do the causal heavy lifting in cases where we seem to be sensitive to differences in the normative properties. I conclude that normative minimalism amounts to a form of epiphenomenalism which is no more plausible than epiphenomenalism about conscious experience.

#### 4.1. *Normative truths and the correspondence theory of truth*

What does it mean to say that normative truths lack positive ontological implications, as opposed to, say, the truths of the natural sciences, which obviously do have ontological implications? One way of explaining this is to appeal to the correspondence theory of truth. Under the correspondence theory, truths have “truthmakers” which explain why they are true, a “truthmaker” for a truth being “an entity in the real world, a thing distinct from the truth itself, with which the truth, as a whole, is correlated” (Dodd 1999, 147). If a proposition is true under the correspondence theory of truth, this implies that some entity exists in the real world, distinct from the truth itself, which makes the truth true, and which is therefore an ontological implication of that truth.

In the worldly sense of “fact,” a truthmaker is a fact, and as Skorupski (2010) points out “*of course* facts can make a normative proposition true. What makes it true that you are acting wrongly is the fact that you are causing suffering, and doing so because of the pleasure it gives you” (434). Is the fact that you are causing suffering because of the pleasure it gives you an entity that exists in the real world, distinct from the normative truth that you are acting wrongly, and therefore one of the truthmakers for a normative truth in the sense intended by the correspondence theory? This fact is probably distinct from the normative truth that you are acting wrongly.<sup>62</sup> But there is a rather trivial way in which plenty of normative truths have ontological truthmakers: if you are acting wrongly because you are causing suffering just for pleasure, then of course this implies that you exist, and that the suffering you are causing exists, and these are both (in some sense) entities,<sup>63</sup> distinct from any truth, which exist in the real world (and which make certain truths, including normative truths, true). This is not something that is at issue between the normative minimalist and the Robust Realist; the minimalist is aware that entities of certain kinds (such as, for instance, moral agents and their actions) often make normative

---

62. Though if some form of reductive normative naturalism is true, the normative fact that you are acting wrongly may be the same fact as the natural fact that you are causing suffering just for pleasure, or (more likely) the same fact as some disjunctive natural fact of which causing suffering just for pleasure is one of the disjuncts.

63. Or “existents,” if “entities” is too ontologically loaded a term.

truths true.<sup>64</sup> What the minimalist should be taken to deny is that *fundamental* normative truths have truthmakers, in the sense of “real” entities that explain why these truths are true. It is not a *fundamental* normative truth that you are acting wrongly because you are causing suffering just for pleasure, because this normative truth depends on a non-normative truth (that you are causing suffering just for pleasure). That it is wrong to cause suffering just for pleasure, however, probably is a fundamental normative truth. For the minimalist, *this* truth, along with other fundamental normative truths, has no positive ontological implications because it neither has nor needs a truthmaker.

One can see why certain truths might need truthmakers, whereas other truths might not. Quine (1979) believed that science has some title to a correspondence theory of truth because of its link to observation, whereas ethics does not, because, in Quine’s view, it lacks this same link (477). If normative truths are not discovered through empirical observation and experiment in the way that scientific truths are,<sup>65</sup> this could indicate that scientific truths have truthmakers whereas (fundamental) normative truths do not. Minimalists like Parfit and Skorupski accept an *a priori*

---

64. Skorupski, after all, is a minimalist, or a “cognitive irrealist,” as he puts it (439).

65. This issue touches heavily on the debate between normative naturalism and normative non-naturalism, so I will not attempt to resolve it in this chapter. In Chapter 7, I will defend an empirical account of normative knowledge according to which the normative domain *does* have a link to observation.

model of normative knowledge,<sup>66</sup> and *a priori* truths are more plausibly construed as lacking ontological implications than empirical truths. Empirical truths, after all, are clearly about entities that exist in the real world, so it is perfectly natural that real entities should serve as truthmakers for these truths. It is less obvious, however, that *a priori* truths are about real entities at all.<sup>67</sup> And if they are not about real entities, their truth is not held hostage to the existence of such entities.<sup>68</sup> If (fundamental) normative truths are *a priori*, therefore, this could explain why they might lack positive ontological implications. The correspondence theory of truth does not seem quite as apt for *a priori* truths as it does for empirical truths.

---

66. We will also look at Parfit's and Skorupski's accounts of normative knowledge in Chapter 7.

67. One could of course argue that abstract objects are no less real than concrete objects just because knowledge of them is *a priori*, but I think that concrete objects are still more obviously real than abstract objects, and that empirical truths about concrete objects are therefore more obviously in need of truthmakers (even if both empirical and *a priori* truths in fact need truthmakers). As Armstrong (2010) says, the spatio-temporal world "seems obviously to exist. Other suggested beings seem much more hypothetical" (1).

68. By way of analogy, Parfit (2017) claims that mathematicians need not fear that all mathematical propositions might be false because numbers don't exist in some Platonic realm, since mathematical truths are *a priori* and therefore (according to the minimalist) lacking in ontological implications (62).

#### 4.2. Normative truths and modal truths

However, if Armstrong's (2004) "Truthmaker Maximalism" is correct, then all truths, including *a priori* truths, have ontological truthmakers (5).<sup>69</sup> Armstrong argues that necessary, *a priori* truths such as  $\langle 7 + 5 = 12 \rangle$ , have truthmakers just like contingent, empirical truths such as the truth that I have five fingers on my right hand (99). If mathematical truths need truthmakers in spite of being *a priori*, then even if (fundamental) normative truths are also *a priori*, this will not automatically exempt them from a general need for truthmakers. And if *modal* truths—truths about what is possible and what is necessary—uniformly possess truthmakers, this could indicate that normative truths have truthmakers as well. Parfit suggests a similarity between normative and modal truths in terms of their (non-)metaphysical status. According to Parfit, "[t]here are ... many events and other things that are merely possible" but "[t]hese merely possible things are not observable features of the spatio-temporal world," and "[w]hen we claim that there are some things we *could* have done, these claims do not commit us to the existence of strange entities as parts of reality" (2011b, 487). In other words, modal truths have no positive ontological implications.

---

69. Armstrong admits that he does not have a direct argument for Truthmaker Maximalism, but his "hope is that philosophers of realist inclinations will be immediately attracted to the idea that a truth, any truth, should depend for its truth on something 'outside' it, in virtue of which it is true" (7). But if he can show that there are some *a priori* truths which do have truthmakers, this will undermine the reason for thinking that fundamental normative truths do not have ontological implications just because they are *a priori* (if indeed they are *a priori*).

Yet Armstrong has a strong argument for the conclusion that modal truths *do* have truthmakers: take a proposition  $p$ —if  $p$  is contingent (i.e. not a necessary truth), then  $p$  entails  $\langle$ it is possible that not- $p$  $\rangle$  which is a modal truth. Now say that  $p$  has a real entity  $T$  as its truthmaker; if  $T$  is the truthmaker for  $p$ , then, since  $T$  explains why  $p$  is true,  $p$  is true in virtue of  $T$ , so, if  $T$  exists, then, necessarily,  $p$  is true. But if  $p$  is true, since  $p$  is contingent, then, necessarily,  $\langle$ it is possible that not- $p$  $\rangle$  is also true. Therefore, if  $T$  exists, then, necessarily,  $\langle$ it is possible that not- $p$  $\rangle$  is true, and this makes  $T$  itself the truthmaker for this modal proposition. The truthmaker for the modal proposition  $\langle$ it is possible that not- $p$  $\rangle$  is the same as the truthmaker ( $T$ ) for the contingent, non-modal proposition  $p$  (2004, 84).<sup>70</sup> The modal proposition  $\langle$ it is possible that not- $p$  $\rangle$  entails that  $T$  (a real entity) exists, and therefore this and other modal propositions do have ontological implications in spite of being necessary and *a priori* (if  $p$  is contingent, then  $\langle$ it is possible that not- $p$  $\rangle$  is a necessary, *a priori* truth).

Does the fact that modal truths like  $\langle$ it is possible that not- $p$  $\rangle$  have truthmakers give us enough reason to believe that fundamental normative truths also have truthmakers? It does give us reason *not* to believe that fundamental normative truths *don't* have truthmakers just because they are *a priori* (if they are *a priori*). But it does not appear to give us any reason to think that normative truths have their own truthmakers *in addition* to the truthmakers that already exist for non-normative truths. Armstrong claims that he has “removed the need for any

---

70. This argument uses what Armstrong calls the “Entailment principle”: “[s]uppose that  $T$  is a truthmaker for proposition  $p$ . Suppose further that  $p$  entails proposition  $q$  ... Then  $T$  will be a truthmaker for  $q$ ” (10).

truthmakers for truths of 'mere possibility' except the truthmakers for contingent truths" (84). And according to his argument, the truthmaker for <it is possible that not- $p$ > is the very same truthmaker as the truthmaker for  $p$ . For all this shows, modal truths have truthmakers only because they are entailed by non-modal truths which have truthmakers. The Entailment principle means that, if fundamental normative truths are entailed by non-normative truths which have truthmakers, then normative truths also have truthmakers, for the same reason that modal truths have truthmakers. But just because modal truths are entailed by non-modal truths does not mean that fundamental normative truths are entailed by non-normative truths. And normative minimalists like Parfit and Scanlon, viewing the normative as a fully autonomous domain, unequivocally reject the notion that fundamental normative truths are ever entailed by non-normative truths.<sup>71</sup>

Fundamental normative truths are probably necessary truths: if it is indeed wrong to cause suffering just for your own pleasure, there is not some other possible world in which this behaviour is somehow not wrong. Armstrong thinks that the truthmakers for necessary truths "are the entities, whatever they are ... which are involved in the truths. These, and these alone, will be the truthmakers for necessary

---

71. Skorupski also believes that "propositions about reason relations are not about any domain of substantial facts [what we are calling worldly facts]" for the simple reason that "factual and normative propositions are different" (2010, 402). He endorses a fundamental "distinction between descriptive and normative ... between propositions about the world and propositions about reasons," factual propositions (in his sense) being propositions about the world and normative propositions being propositions about reasons (4). If this distinction is valid, then no non-normative proposition (a proposition about the world) ever entails a fundamental normative proposition (a proposition about reasons).

truths." For instance, "[t]he necessary truth  $\langle 7 + 5 = 12 \rangle$  has as its truthmakers the natural numbers 7, 5 and 12." If all necessary truths have truthmakers, what kind of entities would serve as truthmakers for normative truths? In the mathematical case, Armstrong thinks that a natural number "is some sort of *property* of things ... presumably a property of certain classes and/or mereological wholes" (2004, 99). If he is correct, then  $\langle 7 + 5 = 12 \rangle$  describes a necessary relation between three properties (7, 5 and 12), a relation that holds entirely in virtue of the nature of the properties themselves—and that is why these properties are the truthmakers for this truth. In the normative case, the necessary truth that  $\langle$ it is wrong to cause suffering just for your own pleasure $\rangle$  would describe a necessary relation between the properties  $\langle$ causing suffering just for your own pleasure $\rangle$  and  $\langle$ being wrong $\rangle$ . Because of the nature of these two properties, if a relevant particular (an action, say) has the first of these properties, then, necessarily, it also has the second. These two properties, therefore, would be the truthmakers for this normative truth: if one or both of these properties did not exist, it would clearly not be true that it is wrong to cause suffering just for pleasure. So this normative truth implies that these properties exist. Doesn't this clearly establish that (fundamental) normative truths have positive ontological implications after all?

#### 4.3. *Two senses of "exist"*

In one sense, normative truths imply that normative properties exist. For the normative minimalist, however, this is not enough to entail that normative truths

have ontological implications. Although he of course accepts the existence of normative properties, Parfit denies that these properties exist “in some ontological sense” (2011b, 479). His position depends on distinguishing between two different senses of “exist”: the ontological and the non-ontological senses. According to Parfit, the ontological sense of

‘exists’ is not ... the only important sense. If nothing had ever existed in any ontological sense, there would not have been any stars or atoms, nor would there have been space, time, or God. But it would not have been true that nothing ever existed ... there *would have been* the truth that nothing existed in an ontological sense. This truth would have existed in a different *non-*ontological sense. (485)

In an empty universe, a universe in which nothing exists, it would be true that nothing exists, in which case, in that universe, there would *be* the truth that nothing exists, and that truth could therefore be said to exist in that universe. Wouldn't the existence of that truth be inconsistent with the emptiness of that universe, meaning that that universe is not in fact empty after all? But in an empty universe, how could it not be true that the universe is empty, and if it is true that the universe is empty, how could that truth fail to exist? One solution to this apparent paradox could be that empty universes are just impossible, because the idea of such a universe involves precisely this contradiction (nothing exists, therefore the truth that nothing exists exists, therefore it is not the case that nothing exists). Another solution—the solution Parfit favours—is that empty universes are of course possible (logically possible, at any rate) and thus the truth that nothing “exists” does itself “exist” in a universe in which nothing “exists,” but this doesn't entail a contradiction because

the sense in which nothing “exists” in an empty universe is the ontological sense, whereas the sense in which the truth “exists” is the non-ontological sense.

Empty universes are at least logically possible: it is not an analytic truth that there are no such universes. Empty universes may in fact be impossible (because some things necessarily exist) but this is not something that follows from the meanings of the words “truth” and/or “exists.” So Parfit is probably right that there are two senses of “exist” such that the truth that nothing exists can consistently “exist” in a universe in which nothing “exists.”<sup>72</sup> But that does not automatically mean that one of these senses is “ontological” whereas the other is “non-ontological.” The distinction between the two senses of “exist” could be a distinction between concrete and abstract existents: stars or atoms, after all, are clearly concrete existents, whereas truths, facts and properties are (arguably) abstract existents. Although it is more difficult to classify things like space, time or God as concrete or abstract, it is also unclear what it means (or would mean) for such things to exist in

---

72. Skorupski provides another example to illustrate this point:

Consider this statement:

- (1) There are characters in *War and Peace* who do not exist and characters who do.

Here ‘exist’ is clearly used as a predicate to make an ontological distinction between the fictional characters and the real people who feature in the novel. However, ‘exists’ can also be used to express the existential quantifier. When ‘exists’ expresses the quantifier, ‘Fs exist’ has the same force as ‘there are Fs’ ... In this use it is trivially true that whatever domains semantics posits exist. (2010, 423)

There is clearly one sense in which fictional characters “exist” (because there *are* such fictional characters) and another sense in which, being fictional, these characters do not exist. There *is* such a fictional character as Sherlock Holmes, so in one sense Sherlock Holmes “exists”; but of course, Sherlock Holmes does not exist.

an “ontological” sense rather than the “non-ontological” sense in which truths are supposed to exist.

Skorupski, whose “cognitive irrealism” (2010, 439) is very similar to Parfit’s “Non-Realist Cognitivism” (2017, 56), attempts to clarify the distinction between the ontological and non-ontological senses of ‘exist’ by appealing to the notion of causal standing: “to exist [in the ontological sense] is to have causal standing ... To be a substantial property, an attribute, is to have causal standing, where a property has causal standing if an object that has the property can cause or be caused in virtue of having it” (2010, 428). And Skorupski holds that normative properties (and relations) are irreal because they have no causal standing (439). Whether or not Skorupski is right that all normativity is fundamentally about reasons,<sup>73</sup> I think that causal standing does explain the difference between things that do and things that don’t have (positive) ontological implications. When Parfit denies that normative properties “exist” in some ontological sense, he should be taken to be denying that such properties have causal standing. This makes sense of the distinction between the way in which things like atoms and planets exist and the way in which things like truths exist: atoms and planets cause things (and are caused themselves), truths do not.<sup>74</sup> Certain properties, such as the physical properties of objects, also cause

---

73. He calls this the “Reasons thesis” (2). Scanlon subscribes to a similar view which he calls “Reasons Fundamentalism”; he too is inclined to believe that reasons are the only fundamental elements of the normative domain (2014, 2).

74. Although facts (in the worldly sense) do cause things, and many facts are themselves caused to obtain. One might think that truths also have causal standing:

things—the properties of the earth and the sun (such as their relative masses) are what cause the earth to orbit the sun (rather than the other way round). For the normative minimalist, normative properties make normative propositions true but they do not cause things, and that is why (fundamental) normative truths do not have ontological implications.

One might wonder why there could not be ontologically weighty properties or entities which have no causal standing at all. If mathematical Platonism is true, wouldn't Platonic numbers "exist" in the ontological sense, but still lack causal standing of any kind? I think this would be the case only if we construe causal standing in a narrow sense, as involving something like energy transfer between concrete spatio-temporal objects. Of course, Platonic numbers would not have causal standing in this sense. If, on the other hand, we construe causal standing more broadly as covering any kind of productive power (Skorupski 2010, 426-427), the power to produce or influence contingent circumstances, then there is no problem with interpreting the distinction between ontological and non-ontological existence in terms of causal standing. Under mathematical Platonism, we acquire mathematical knowledge by some form of interaction with an independent mathematical realm, a realm which has some form of power to produce

---

the truth of a proposition, we might say, makes it more likely that it will be believed by rational creatures like us, and in cases where we do believe the proposition, the truth of this proposition is one of the things that causes us to believe it. But truths, of course, are propositions, and I am inclined to think that it is the (worldly) facts and properties that serve as truthmakers for propositions which cause things (such as our beliefs), and not the propositions themselves.

mathematical beliefs (and knowledge) in us. Similarly, if a robust version of metanormative realism is true, then at least some normative properties must have the power to produce certain beliefs in us, and to influence us to behave in certain ways. If normative properties have some kind of causal standing, these properties exist in the ontological sense, and the normative truths that involve these properties have (positive) ontological implications as a result. And this, I believe, is the very heart of the issue.

#### 4.4. *Minimalism and fictionalism*

Given that there are two senses of “exist,” one of which (the “ontological” sense) implies causal standing, whereas the other implies the absence of causal standing, normative minimalism entails that normative facts and properties have no causal standing—that they “exist” only in the (“non-ontological”) sense that does not imply causal standing. Enoch rejects minimalism on the grounds that it is essentially a notational variation of normative *fictionalism*,<sup>75</sup> i.e. the non-realist view according to which normative truths are only true, and normative facts and properties only exist, *within the fiction* of normativity, in the same way that truths about Sherlock Holmes (such as the truth that he smokes a pipe) are only true, and Sherlock Holmes himself only exists, within the Sherlock Holmes fiction (2013, 124). After all, it is clear that Sherlock Holmes does not exist in the spatio-temporal world, and that he does not exist in some non-spatio-temporal part of reality either, but it is still true that

---

75. Like McPherson, he refers to minimalism as “quietism” (2013, 121).

Sherlock Holmes smokes a pipe, and as we have seen, a minimalist like Parfit would say that, although a normative property like moral wrongness does not exist in the spatio-temporal world, or in some non-spatio-temporal part of reality, it is still true that it is morally wrong to cause suffering just for pleasure. Sherlock Holmes does exist in some sense,<sup>76</sup> but he does not exist in the ontological sense, the sense which implies causal standing, and neither, for the minimalist, do normative facts and properties; if normative facts and properties only “exist” in the same sense in which a fictional character like Sherlock Holmes “exists,” one might think that truths involving these facts and properties (normative truths), just like truths about Sherlock Holmes, cannot be mind-independent.<sup>77</sup> Assuming, as I argued in the previous chapter (Section 3.5), that mind-independent normative truths are deliberately indispensable, normative fictionalism must be false, so normative minimalism, if it cannot be substantively distinguished from fictionalism, must be false as well.

The minimalist believes that first-order normative claims are licensed entirely by standards internal to the normative domain and normative discourse, as are claims to the effect that normative properties and facts exist, and that normative reasons exist. For this reason, Enoch argues that the minimalist is also committed to the existence of *counter*-reasons, first-order claims about which are licensed by standards internal to a *counter-normative* discourse: a discourse that is similar to

---

76. Recall Skorupski’s point from footnote 72.

77. It is not a mind-independent truth that Sherlock Holmes smokes a pipe.

normative discourse in so far as those engaged in this discourse “treat counter-reasons much as we treat reasons ... they take them to be relevant to their practical deliberation, or perhaps counter-deliberation, in much the same way we take reasons to be relevant to ours” (124).<sup>78</sup> And this presents an insoluble problem for the minimalist, according to Enoch, because “when deliberating, one of the questions you may ask yourself is whether to ‘go’ with reasons or with counter-reasons,” and the minimalist “has nothing by way of an adequate reply” (126). In other words, if reasons and counter-reasons both exist, they are both relevant to deliberation, and they both recommend completely different decisions, then we have no non-arbitrary way of deciding whether to base our decisions on reasons or on counter-reasons, and hence no non-arbitrary way of arriving at any deliberative decision.

Scanlon regards this problem as an illusion, because the conclusions about counter-reasons reached by participants in the counter-normative discourse “conflict with our conclusions about reasons only insofar as they are interpreted as conclusions about reasons. So the question of which one of us is correct is a normative question, which can be answered only through normative reasoning” (2014, 29). In other words, for Scanlon, there can be no question about whether to “go” with reasons or counter-reasons in deliberation, because either counter-reasons are relevant to deliberation, in which case they are just reasons, or they are not

---

78. Counter-reasons are supposed to be the kind of considerations that we would not consider as reasons. Enoch stipulates that those who engage in counter-normative discourse “think that it is rather obvious that that an action will cause the agent pain is counter-reason *for* performing it” (125).

reasons, in which case they are not relevant to deliberation. The problem is supposed to be generated by the incommensurability of reasons and counter-reasons—they belong to different domains or discourses—plus the equal relevance of both reasons and counter-reasons to deliberation (we might just as well “go” with counter-reasons in deliberation as with reasons). But if reasons and counter-reasons are both equally relevant to deliberation, then, far from being incommensurable, they are in fact part of the very same domain, that is, the *deliberative* domain, and hence they are both considerations—*deliberative* considerations—of the same fundamental kind. And since the kind of deliberation in question is clearly the normative kind, the deliberative domain is just the normative domain. As far as Scanlon is concerned, therefore, the problem Enoch tries to create for the normative minimalist is incoherent, and his argument against minimalism fails.

Counter-reasons may indeed be just reasons, as Scanlon says, but this does not eliminate the problem suggested by Enoch’s analogy between minimalism and fictionalism. In Chapter 2 (Section 2.4), we saw that normative truths cannot be entirely relative to frameworks: the normative domain cannot be relativistic all the way down.<sup>79</sup> But this is exactly the scenario that would obtain if the normative domain is akin to a fiction like the Sherlock Holmes fiction, the truths of which are only true within that particular fiction and not outside. If it is true that it is wrong to

---

79. Unless, that is, normative deliberation is incapable of arriving at anything but (normatively) arbitrary answers, in which case we would have no normative reason to engage in it. By Parfit’s Wager, however, we do have normative reason to engage in normative deliberation.

cause suffering just for pleasure only within the normative fiction, just like it is true that Sherlock Holmes smokes a pipe only within the Sherlock Holmes fiction, then causing suffering just for pleasure is clearly only wrong relative to a particular framework, and might just as well not be wrong relative to some other framework, in the same way that Sherlock Holmes might smoke a pipe in one fiction but not in another.<sup>80</sup> Unless Sherlock Holmes exists in the ontological sense, there is no fact of the matter as to whether or not Sherlock Holmes smokes a pipe, and since Sherlock Holmes does not exist in that sense, the “truth” that he smokes a pipe is not a mind-independent truth. For the minimalist, normative properties and facts do not exist in this same sense, yet somehow this does not prevent the normative domain, unlike the Sherlock Holmes fiction, from containing determinate, mind-independent truths. So what, in that case, differentiates the normative domain from a piece of fiction like the Sherlock Holmes fiction?

The Robust Realist can answer that normative properties and facts exist in the ontological sense (that they have some kind of causal standing) whereas Sherlock Holmes does not, and this explains why there are some fully determinate, mind-independent normative truths and yet there are no such truths about Sherlock Holmes. The minimalist of course cannot say this. What the minimalist can say, however, is that fundamental normative truths, unlike truths about Sherlock

---

80. Compare the original Conan Doyle stories with the modern BBC adaptation (in which he doesn't smoke a pipe). These might both be described as variants of the Sherlock Holmes fiction, just like, under normative fictionalism, different normative frameworks, to which different normative truths are relative, might be described as variants of the normative fiction.

Holmes, are *a priori*<sup>81</sup>; if it is an *a priori* truth that it is wrong to cause suffering just for pleasure, this normative truth will not be true only relative to a certain fiction or framework, and that is so regardless of whether or not the normative property of being wrong has some kind of causal standing. The minimalist can also say that fundamental truths about reasons are knowable *a priori*, whereas there are no such *a priori* truths about counter-reasons, even if some first-order claims about counter-reasons are licensed by the internal standards of a counter-normative domain<sup>82</sup>; the participants in the counter-normative discourse have no *a priori* knowledge about counter-reasons, whereas we, as participants in the normative discourse, do have *a priori* knowledge about reasons. Since the normative minimalist can differentiate the normative domain from merely fictional domains by appealing to *a prioricity*, the Robust Realist must argue more directly for the conclusion that normative properties have a kind of causal standing.

---

81. For Skorupski, normative reason relations and fictional objects are both unreal (they have no causal standing), but what distinguishes mind-independent reason relations from mind-*dependent* fictional objects is that reason relations, unlike fictional objects, are “objects not of imagination but of *pure cognition*” (2010, 429). I take it that pure or fundamental truths about objects of “*pure cognition*” are known *a priori*.

82. Scanlon would presumably say that it is an *a priori* truth that that an action will cause the agent pain is a reason against performing it, whereas the equivalent counter-reasons claim—that that an action will cause the agent pain is a counter-reason *for* performing it—is not *a priori*, although it may be true relative to the counter-reasons framework.

#### 4.5. *Do normative properties have causal standing?*

Larry Temkin (2016) argues that “on both major philosophical accounts of causation—the counterfactual model and the deductive-nomological model—reasons would count as having a causal impact on our actions,” and therefore that “reasons, or normative facts, can have a causal impact on us and, a fortiori, that they can have a causal impact, though us, on the world” (3-4). On the counterfactual model of causation, “A is a cause of B whenever it is the case that ‘but for’ A, B wouldn’t have occurred” (4). According to Temkin, if, for instance, it is a normative fact that I ought to stop my car at a red light, and, recognizing and responding appropriately to this normative fact, I stop at the red light, then, if, for whatever reason, it had instead been the case that I ought *not* to have stopped at the red light, since I am recognizing and responding appropriately to the normative facts, I would *not* have stopped at the red light (5). And so, “but for” the normative facts being as they are and not otherwise (it being the case that I *ought*, rather than *ought not*, to stop at the red light), I would not have done what I actually did (stopping at the red light), meaning that my action is counterfactually dependent on the normative facts, and thus that these normative facts, and the normative properties involved in them—by the counterfactual theory of causation—must have some kind of causal standing.

The causal efficacy of normative facts and properties seems to be indicated by a fact that Parfit himself not only acknowledges but emphasises as integral to his position, namely, the fact that we have the capacity to respond appropriately and

sensitively to normative reasons, and hence (it would seem) to the normative facts and properties themselves. In Parfit's own words, "when it is true that ... we have decisive reasons to act in some way, this fact makes it true that ... if we were fully informed and both procedurally and substantially rational, we would choose to act in this way" (2011a, 63). Say that we do have decisive reason to act in some way, and we are fully informed and both procedurally and substantially rational, and so we do, therefore, act in that way; does this not imply that, had we been just as informed and rational as we in fact are, but, instead of having decisive reason to act in that way, we actually had decisive reason to act in a *different* way, we would have acted in that different way, and thus that our acting in the way we actually do is counterfactually dependent on the normative fact that we have decisive reason to act in that particular way? In other words, if we could, in principle, hold all other variables fixed, and just change the normative facts sufficiently, would this not produce a different result in terms of our actions? And since our actions affect the world, does this not entail that normative facts, through our actions, affect the world as well (and therefore that they have some kind of causal standing)?

I think Parfit would answer that normative reasons do affect the world through our responses to them: if the fact that I am at a red light is a reason for me to stop my car, and I respond to this reason I have by stopping my car, then one of the things that causes me to stop is clearly the fact that I am at a red light. And since this fact is also a reason, normative reasons are able to affect the world causally just like natural facts, because, after all, reasons *are* natural facts. Yet just because normative

*reasons* are causally efficacious does not entail that normative *facts* and *properties* have any causal standing of their own independently of the purely natural, non-normative facts and properties which undoubtedly cause our actions and affect the world. Even though the natural, non-normative fact that I am at a red light is no doubt one of the things that causes me to stop my car, this does not entail that the specifically normative fact that I ought or have reason to stop at the red light also plays a role in causing my action. But in spite of this lack of entailment from the causal efficacy of the natural, non-normative fact (that I am at a red light) to the causal efficacy of the normative fact (that I ought to stop at the red light), this normative fact definitely does seem to play a causal role of its own: had it been the case that I ought *not* to have stopped at the right light, then, assuming that I am just as responsive to the normative facts and properties in this counterfactual scenario as I am in the actual world, I surely would *not* have stopped at the red light. It follows from this that, if the normative facts had been different, and if I am fully informed and both procedurally and substantially rational (which I am by the hypothesis), I would have acted differently than I actually did as a result. And thus my actions, at least some of the time, appear to be counterfactually dependent, not just on the normative reasons and the non-normative facts which constitute them, but on the specifically normative facts and properties as well, suggesting that these facts and properties do indeed have some kind of causal standing.

How might Parfit respond to this pretty straightforward argument? As far as I am aware, Parfit never responded to Temkin,<sup>83</sup> but if he had, I believe he would have argued that our actions cannot be counterfactually dependent on the normative facts, because normative truths and normative facts are necessary, and facts that obtain necessarily cannot support counterfactuals.<sup>84</sup> We cannot say truly, Parfit might argue, that “but for” a certain normative fact, I would not have acted as I did in any given situation, because this implies that there is a possible world—some alternative counterfactual scenario—in which this normative fact did not obtain, and in which I therefore acted differently than I did in the actual world as a result. But if normative facts are uniformly necessary, there simply is no such possible world—no alternative counterfactual scenario—and this counterfactual therefore cannot hold. After all, the mathematical fact that  $\langle 7 + 5 = 12 \rangle$  is quite plausibly regarded as causally inert: nothing counterfactually depends on this mathematical fact, because it obtains necessarily, and there is thus no possible world in which anything is different because this fact does not obtain in that world. If normative facts are necessary, just like the fact that  $\langle 7 + 5 = 12 \rangle$ , then we could not, even in principle, change the normative facts, while holding all other variables constant, to produce a

---

83. In spite of dedicating his *Volume Three of On What Matters* (partly) to him.

84. For Parfit, since normative truths do not have ontological implications and do not have truthmakers, there are normative facts only in the propositional sense of “fact” (i.e. a normative fact is nothing but a true normative proposition), and so the normative facts and the normative truths are all one and the same.

different result in our actions and in the world, so normative facts cannot feature in counterfactual relations and hence cannot have causal standing of any kind.

On the face of it, this response I have made on Parfit's behalf is easy to dismiss on the grounds that, while some normative facts may be necessary and thus unable to support counterfactuals, some normative facts are certainly contingent, and our actions therefore *can* be counterfactually dependent on these facts. It is clearly false that, if it is a fact that I ought to stop at a red light, this normative fact obtains necessarily: circumstances could have been such that I very much ought not to stop (if I desperately needed to get to a hospital as quickly as possible, say), in which case, if I had been just as responsive to the normative facts as I in fact am, I would not have stopped. In such cases, there seems to be no reason to deny that my action is counterfactually dependent on the normative facts. Granted, if it is indeed morally wrong to cause suffering just for pleasure, this is probably a necessary truth, such that there is no possible world in which anyone acts differently than they do in the actual world because it is not wrong to cause suffering just for pleasure (in that world). But just because there are some necessary normative facts which cannot support counterfactuals does not mean that no normative facts can or that normative facts as a whole do not support counterfactuals.

But now suppose that fundamental normative facts are indeed necessary, and that facts which obtain necessarily cannot support counterfactuals. In this scenario, fundamental normative facts would have no causal standing, and fundamental normative truths would have no ontological implications. Non-fundamental or

mixed normative facts (such as the fact that one ought to stop at a red light under specific non-normative circumstances) may be contingent and thus able to support counterfactuals, and mixed normative truths may have ontological implications as a consequence, but a mixed normative truth is nothing but the combination of a fundamental normative truth and a non-normative truth, and since, by the hypothesis, the mixed normative truth cannot derive its ontological implications from the fundamental normative truth (such truths having no ontological implications), the ontological implications it does have must come entirely from the non-normative truth. From this it follows that the only normative truths which have ontological implications (i.e. mixed normative truths) do not have any ontological implications *of their own*, and in fact only have ontological implications at all because they are partly non-normative truths. Unless either fundamental normative truths are not necessary or necessary truths are able to support counterfactuals after all, normative truths *qua* normative truths do not have ontological implications (and normative facts and properties do not have causal standing). Or so, I suggest, Parfit and other normative minimalists would argue.

As we noted in Section 4.1, the normative minimalist should be taken to deny only that *fundamental* normative truths and normative facts have ontological implications/truthmakers and causal standing, not that mixed normative truths/facts do, because mixed normative truths/facts are partly composed of non-normative truths/facts that clearly do have ontological implications and causal standing. If fundamental normative truths/facts are necessary and hence unable to support

counterfactuals, this provides some reason to think that such facts are not causal or ontological in nature.

Yet I think we still have reason to believe in ontological normative facts. Even if fundamental normative facts obtain necessarily, such facts will only fail to have ontological implications themselves if the normative properties featured in these facts also lack any kind of causal standing: if a necessary truth is about properties that have causal standing, these properties will be ontological truthmakers for this truth, and this truth will have ontological implications through these properties even though the truth itself is unable to support counterfactuals. And just because certain properties feature in necessary truths does not mean that these properties lack any kind of causal standing. The mathematical truth that  $\langle 7 + 5 = 12 \rangle$  is clearly a necessary truth, and the properties involved in this truth probably do have a kind of causal standing: if the fingers on my right hand did not instantiate the property of being five, I would not see what I actually see when I look at my right hand. Mathematical properties (such as the property of being five) do have the power to produce and influence contingent, non-mathematical circumstances (such as by making me have a certain visual experience when I look at my right hand). That a mathematical truth like  $\langle 7 + 5 = 12 \rangle$  is necessary does not prevent the mathematical properties involved in this truth from having causal standing. And since mathematical properties do appear to have causal standing,<sup>85</sup> and the necessary

---

85. Some mathematical properties, such as the property of being prime, probably don't have any causal standing: everything that has the property of being

truths of arithmetic are true entirely in virtue of these properties and the relations holding between them, we can say that these properties constitute ontological truthmakers for these truths, and that these truths, even though they are necessary, therefore have positive ontological implications. Can't we say the same thing about normative properties, even if fundamental normative truths are uniformly necessary?

Parfit insists that "[w]hen some fact has the property of being or giving us a reason, we cannot be causally affected by this normative property." He thinks that our responsiveness to reasons involves no causal contact with normative properties. But although this responsiveness may not be causal in some narrow sense, involving spatio-temporal energy transfer, or direct object-to-object physical contact or the like, it is difficult to understand how it could not be causal in the broad sense in which causality simply involves productive or influential power of any kind (productive or influential power itself being understood in counterfactual terms). Fundamental normative truths may not support counterfactuals, but normative properties clearly do: if different non-normative facts instantiated the property of being a reason, then,

---

prime (the prime numbers) has this property necessarily, meaning that this property does not support counterfactuals. Yet the same does not appear to be the case for the natural numbers, considered as properties of classes and/or mereological wholes. And the point remains that just because such properties are involved in necessary truths (which do not support counterfactuals) does not mean that these properties do not support counterfactuals themselves. The fact that fundamental mathematical truths (truths which have no non-mathematical content) are necessary does not prevent the mathematical properties involved in them from having causal standing. Even if fundamental normative truths are necessary, therefore, this should not prevent the normative properties involved in these truths from having causal standing either.

assuming that we are sometimes responsive to reasons, we would sometimes respond and act differently as a result. And even if certain properties have causal standing only in the broad sense, this is surely enough for the truths involving these properties to have ontological implications: if God exists beyond space, time and causality (in the narrow sense), then the truth that He created the universe would still have ontological implications. Parfit himself says that “we respond to reasons when we are aware of facts that give us these reasons, and this awareness *leads us* [my emphasis] to believe, or want, or do, what these facts give us reasons to believe, or want, or do” (2011b, 493). If our awareness of the reasons we have leads us to believe what we have reason to believe, or to do what we have reason to do, then our reasons themselves, by means of our awareness of them, lead us to respond in these ways, and this surely suffices for the property of being a reason to have causal standing (at least in the broad sense).

Parfit points out that our mental abilities allow us to form true mathematical beliefs by reasoning in valid ways, and he claims that, “[s]ince we cannot be causally affected by the validity of these kinds of reasoning, our abilities involve what we can call a *non-causal* response to this validity.” Our capacity to respond to normative reasons and normative properties, he suggests, is analogous to our capacity to respond to valid reasoning: both kinds of responses are non-causal, yet not somehow miraculous or *contra-causal* (500). But since, if a certain piece of reasoning is indeed valid (in the sense that its conclusion deductively follows from its premises), there is no possibility that this very same piece of reasoning might not

have been valid (because valid reasoning is always necessarily valid), our responses to valid reasoning are not in fact analogous to the majority of cases in which we respond to normative reasons and normative properties. We cannot say truly that, if a certain piece of reasoning, which actually *is* valid, had *not* been valid, we would not have come to form the true beliefs we actually formed on its basis, but we very much can say that, if the fact that I am at a red light was not a normative reason for me to stop my car, then (under certain non-normative circumstances) I would not stop my car. Validity is never contingent upon circumstances, but many of our normative reasons are; differences in our circumstances produce differences in our reasons, which in turn produce differences in our responses. And this appears to be a kind of causal standing.

#### 4.6. *An appeal to supervenience*

Is there any way to resist the conclusion that normative properties have causal standing (of some kind) because they support counterfactuals? It seems undeniable that normative properties support counterfactuals if these properties exist at all and we are genuinely responsive to them. But the minimalist can potentially respond that, since, on any plausible view, normative properties *supervene* on non-normative, natural properties, it is actually these natural properties that do the causal work in cases where non-normative circumstances seem to be

sensitive to differences in the normative properties.<sup>86</sup> Even though I would not stop my car if the fact that I am at a red light did not instantiate the normative property of being a reason for me to stop my car, it is not this normative property, the minimalist could say, that in any way causes me to stop: the natural facts and properties that explain why being at a red light is a reason for me to stop are the things that do the causing. The purely natural fact that stopping at red lights is required by traffic law may explain why the fact that I am at a red light is a reason for me to stop, in which case, it might be this non-normative fact,<sup>87</sup> and not the normative reason it gives me, which actually causes me to stop. For the minimalist who makes this appeal to supervenience,<sup>88</sup> our apparently causal responses to normative properties are really only caused by the underlying non-normative facts that give rise to the instantiation of these properties, and not the normative properties themselves.

One thing to observe here is that, if this Supervenience thesis is combined with the equally plausible *Grounding* thesis, it entails that normative properties are *necessarily co-extensive* with natural properties. Grounding states that “it is impossible

---

86. As Wedgwood (2007) puts it, in the context of the debate about whether or not there are correct “moral explanations” for contingent, non-moral facts: “the question arises whether the alleged effects of a moral fact are really the effects of the moral fact or instead of some non-moral fact in which the moral fact is realized” (193). I will argue for the correctness of certain normative explanations for non-normative facts in Chapter 7.

87. Non-normative, that is, on the reason-involving conception of normativity. This fact may be normative on the rule-involving conception, but this is not the conception that is relevant here (see Chapter 1, Section 1.1).

88. Parfit himself does not make this appeal, instead relying on the analogy with validity that we rejected above.

to have an ethical [or normative] property without also having a non-ethical [or non-normative] property,” and Supervenience states that “there is no difference in ethical [or normative] nature without a difference in non-ethical [or non-normative] nature” (Jackson 2016, 201). It follows that whenever something has certain normative properties, it also necessarily has certain natural properties in virtue of which it has those normative properties. And some take this necessary co-extension between normative and natural properties to indicate that these properties are actually identical to each other. Frank Jackson, for instance, believes that “the necessary co-extension of ethical and natural properties provides a strong reason to identify them” (201).<sup>89</sup> If normative properties really are identical to the natural properties on which they supervene, they will of course have the exact same causal powers and effects as these natural properties. Normative properties will not then have *irreducibly normative* ontological implications, it is true, but, by Leibniz’s Law, they will still qualify as ontologically weighty properties by virtue of being identical to natural properties which obviously do have causal standing in their own right. Supervenience will not save ontological minimalism about the normative if properties that are necessarily co-extensive are also necessarily identical.<sup>90</sup> The normative minimalist, therefore, must maintain that normative properties are

---

89. “Strong but not apodictic,” he adds. Jackson uses the word “natural” in the sense of “non-normative” (or “non-ethical”) (198).

90. I will argue that the necessary co-extension of normative and non-normative properties does not actually indicate that they are identical in Chapter 8 (Section 8.3).

completely distinct from natural properties in spite of being necessarily co-extensive with them.

The necessity of this response on the part of the minimalist—that normative properties do not exert any causal influence of their own, and nor do they reduce to the natural properties on which they nonetheless supervene—shows that normative minimalism essentially amounts to a kind of epiphenomenalism about the normative; and that it is fundamentally implausible for the same reason that epiphenomenalism about conscious experience is implausible. Epiphenomenalism about conscious experience implies that we have many conscious experiences, such as pain, which supervene on physical states of the brain, but which do not reduce to these physical states, and which have no causal influence whatsoever on the physical world and hence no influence whatsoever on our actions: if I instantly retract my hand after placing it on a hot stove, this action has nothing to do with the phenomenal quality of the pain I experience when I place my hand on the stove. Similarly, under normative minimalism, certain facts are often normative reasons for us to do and to believe certain things, but the normative properties of these facts *never* have any influence at all on our actions and beliefs; even when we do and believe exactly what we have normative reasons to do and believe, this is never *because* the relevant facts have the normative property of being a reason. Yet if there are such things as phenomenal conscious experiences, we surely act based on the conscious experiences that we have, such that if we had different experiences—if our experiences had different phenomenal properties—we would sometimes act

differently, and this means that our conscious experiences have some kind of causal standing. And if we have normative reasons for action and belief, as the minimalist believes, then we surely sometimes act and believe based on these reasons, such that if we had different reasons, we would (sometimes) act and believe differently as a result. Since we can act and believe for normative reasons, responding to changes in our normative reasons with changes in our actions and beliefs, the property of being a reason must be capable of having an influence on the real world through our actions and beliefs. Like phenomenal properties, therefore, normative properties seem to possess some kind of causal standing, whether or not they can be reduced to physical or natural properties.

#### 4.7. *Conclusion*

Even if fundamental normative truths are necessary truths, it would seem that normative truths as a whole do have positive ontological implications, in the sense that what makes normative propositions true is the existence of mind-independent entities (normative properties) that have genuine causal standing, at least in the broad sense of that notion. Differences in the instantiation of normative properties make a difference in the real world—the world of contingent facts and circumstances—in the same way that differences in the instantiation of certain mathematical properties, such as the natural numbers, make a difference in the real world, even though the fundamental mathematical truths which involve these properties are necessary and knowable *a priori*. Since we have the capacity to respond to reasons by

(sometimes) forming rational beliefs and by performing rational actions, the property of being a reason must have the capacity to affect us in some way. We should therefore reject ontological minimalism about the normative and opt instead for Robust Metanormative Realism. If we combine this thesis with Occam's Razor, some version of normative naturalism appears to be called for. And this is what we will begin to consider in the next chapter.

## 5. Queerness

If normative facts and properties are not natural facts and properties, one might think that they would have to be, in the words of J. L. Mackie (1990), “utterly different from anything else in the universe” (38) and therefore too “queer” for us to countenance, as he famously put it. One might also think that there are simply unlikely to be any facts and properties that are *utterly* different from anything else in the universe: of course there are many facts and properties that are very different from each other, but all the facts and properties that there actually are are part of one integrated, coherent universe, and thus no real facts and properties differ from each other in such a fundamental way as non-natural normative properties would differ from all other facts and properties if such facts and properties did indeed exist. Given that normative facts and properties exist, therefore, these facts and properties must be natural facts and properties that do not differ fundamentally from the kinds of facts and properties posited by natural science. This is the Argument from Queerness against normative non-naturalism (and when combined with arguments for metanormative realism, such as those I defended in Chapters 2 and 3, *in favour* of normative naturalism).<sup>91</sup>

---

91. Copp (2007) takes the Argument from Queerness to be an argument *against* normative naturalism (i.e. normative properties cannot be natural properties precisely because they are queer and unlike anything else in the universe) (251). I

*Prima facie*, the Argument from Queerness seems to have some considerable force. If we know that there are facts and properties of a certain kind, it is, I believe, by default more likely that facts and properties of this kind are at least not *fundamentally* different from all other facts and properties. But before we dismiss normative non-naturalism as untenable, we must ask exactly why normative facts and properties would be queer if they were non-natural. There could be certain facts and properties that *are* fundamentally different from all other facts and properties in the universe, even if this seems *prima facie* unlikely. Of course non-natural properties are going to be different from all other properties if these other properties are understood as natural properties. This straightforward observation is hardly an argument against non-naturalism about properties of any kind. The whole point of normative non-naturalism is that there is a fundamental “distinction between descriptive and normative ... between propositions about the world and propositions about reasons” (Skorupski 2010, 4). So normative properties, for the non-naturalist, *are* utterly different from all descriptive (i.e. non-normative) properties, but then descriptive properties are utterly different from all normative properties, and neither kind of properties are to be dismissed as queer on that account.

---

don't think it is plausible to take this argument as an argument in favour of non-naturalism: that normative facts and properties would be queer is clearly not an argument for the existence of non-natural normative facts and properties. In defending the Argument from Queerness, Olson (2014) explicitly takes himself to be arguing against non-naturalism (17). And this is how I will understand the argument.

Olson outlines four possible ways of explaining why non-natural normative properties are queer, in the form of “four distinct queerness arguments, focusing on supervenience, knowledge, motivation, and irreducible normativity.” He takes the first three arguments to be failures, but the fourth to be a success (2014, 79). In Sections 5.1 and 5.2, I will argue that Olson moves too quickly in his rejection of the first two queerness arguments (focusing on supervenience and knowledge respectively); he does not consider the possibility of a complete scepticism about *a priori* knowledge (or synthetic *a priori* knowledge at least), a doctrine which is obviously incompatible with normative non-naturalism. As such, I will consider (and reject) the case for this kind of scepticism in Section 5.3, focusing in particular on Quine’s (and Harman’s) arguments against the *a priori* and Bonjour’s responses to them. Then, in Section 5.4, I will address Olson’s argument from the queerness of irreducible normativity: this argument only creates an impasse which Olson illegitimately assumes it is the task of the non-naturalist to resolve. Finally, in Section 5.5, I will draw both upon the conclusion of the previous chapter and an analogy between non-naturalism and mind-body dualism to motivate a further queerness argument which locates the queerness in the interaction between the natural (and non-normative) and the non-natural (and irreducibly normative). The idea that physics is causally closed creates a problem for the non-naturalist, but the non-naturalist can escape this problem by appealing to the causal overdetermination of events. I therefore conclude that the force of the Argument from Queerness essentially reduces to the force of parsimony and Occam’s Razor.

### 5.1. *Queerness and supervenience*

The *first* queerness argument observes that the normative supervenes on the natural, and then states that, if normative properties are distinct from natural properties, there would have to be a unique, non-conceptual form of normative necessity to explain why any particular normative property supervenes on a particular natural property (such as, perhaps, the property of being right and the property of maximizing happiness), and concludes that this relation of normative necessity is queer (90-91). However, the non-naturalist can appeal to some fundamental normative principle(s), like, perhaps, “the property of maximizing happiness is the only property that makes actions right,” to explain the supervenience of the normative on the non-normative, and he can claim that the normative necessity involved in this supervenience is just the necessary truth of such a principle, in which case, the relation of normative necessity will be queer only if non-natural properties and facts are queer (98-99). Olson therefore rejects this first argument because “the charge that ... non-naturalists must take supervenience relations between distinct properties to be brute relations reduces to a general worry about *sui generis* non-natural properties and facts” (100). In other words, it basically amounts to asserting the queerness of non-natural (normative) properties and facts.<sup>92</sup>

---

92. Not to mention that some non-naturalists, such as Terrence Cuneo and Russ Shafer-Landau (2014), think that the fundamental normative principles in question are conceptual truths, and thus that a unique, non-conceptual form of necessity is not needed to explain the supervenience of the normative on the natural.

However, as we noted in the previous chapter (Section 4.6), the supervenience of the normative on the natural means that normative properties are necessarily co-extensive with natural properties, and the necessary co-extension of normative properties with natural properties may indicate that normative properties are identical to natural properties (and are therefore natural properties themselves). Non-natural normative properties would have to be properties that are necessarily co-extensive with natural properties but completely distinct from them at the same time. Streumer's (2017) Reduction Argument purports to show that, "if there are normative properties, these properties are identical to descriptive properties" (40), on the grounds that normative properties are necessarily co-extensive with descriptive properties and necessarily co-extensive properties are necessarily identical. And this argument obviously implies that non-natural, irreducibly normative properties are queer in light of the supervenience of normative properties on natural (or descriptive) properties. I will address the Reduction Argument in Chapter 8 when I explicitly defend non-reductive normative naturalism. So for now I will simply assume with Olson that supervenience is not a problem for the non-naturalist.

---

Their "moral fixed points" are fundamental moral principles that "are true in virtue of the nature of the nonnatural moral [and hence normative] concepts that constitute them" (403). We will consider the plausibility of this claim in Chapter 7 (Section 7.4).

## 5.2. Queerness and the (synthetic) *a priori*

The *second* queerness argument states that, under normative non-naturalism, (fundamental) normative knowledge would have to be synthetic *a priori* knowledge, and synthetic *a priori* knowledge is queer (Olson 2014, 101-102), so normative non-naturalism must be false.<sup>93</sup> Olson responds that this argument generalizes beyond the normative in problematic ways; after all,

Much philosophical knowledge seems to be synthetic *a priori*. Consider, for example, the question of whether there are abstract entities or whether everything that exists is ontologically continuous with a naturalistic view of the world. Knowledge of such matters would seem to be synthetic and *a priori*. (102)

If the naturalist knows that naturalism is true, this very knowledge, one might argue, would itself have to be an instance of synthetic *a priori* knowledge, making it self-defeating for the naturalist to reject normative non-naturalism on the grounds that synthetic *a priori* knowledge is queer. And logical truths like *modus ponens* seem to be both synthetic and *a priori*, suggesting that there is such a thing as synthetic *a priori* knowledge and thus that synthetic *a priori* normative knowledge is (in principle at least) perfectly possible.

Yet naturalism might be best construed, not as a proposed instance of synthetic *a priori* knowledge itself, but as an empirical hypothesis about the nature of the universe and the contingent parameters of our knowledge, to the effect that the

---

93. Assuming, that is, that normative knowledge does exist. Olson does not assume this, because he is considering these arguments in the context of defending the moral (and indeed the normative) error theory. But I am taking myself (by the arguments of Chapters 2 and 3) to have established that normative scepticism is false and that there is such a thing as normative knowledge.

only kind of substantive knowledge that beings like us, in the kind of world that we live in, are capable of possessing is empirical knowledge. The normative non-naturalist is undoubtedly committed to substantive *a priori* knowledge, so if the naturalist has an argument against the existence of this kind of knowledge, he will automatically have an argument against normative non-naturalism. Olson is too quick to dismiss the second queerness argument just because it generalizes beyond the normative. If the naturalist can show that *a priori* knowledge either does not exist at all, or that, if it does, “it is nonetheless merely *analytic* in character—that is, merely a product of human concepts, meanings, definitions, or linguistic conventions” (BonJour 1998, 28), he will also have succeeded in showing that (normative) non-naturalism is false.

### 5.3. *Scepticism about the a priori*

Quinean naturalized epistemology jettisons the whole idea of *a priori* knowledge (analytic or synthetic). Quine (1980) argues that there is no such thing as *a priori* knowledge on the grounds that “no statement is immune to revision,” because any belief can be given up in the face of recalcitrant experience (43). Genuine *a priori* truths would have to be necessarily true statements which, it would appear, could not conceivably fail to hold. But “Quine claims that we can conceive of any statement failing to hold and that for any view we could imagine circumstances in which we would give it up,” and he therefore “concludes that no truths are *a priori* or necessary” (Harman 1967, 132). We probably should accept that all statements are,

on some basic level, open to revision: what seems to be *a priori* knowledge—the Euclidean parallels postulate, for instance—can sometimes turn out not to be true in light of new evidence. Yet even if this is granted, Quine’s conclusion does not follow. Laurence Bonjour points out that “there is no clear reason ... why a proponent of *a priori* justification cannot admit or even insist that such justification is in fact both fallible and corrigible” (1998, 75). Just because a statement or proposition is, *in principle*, open to revision (perhaps in light of experience), does not mean that it is not *in fact* an *a priori* truth; what is needed for the *a priori* is not that there is anything we can know with some kind of absolute certainty, but only that experience is not the only source of (epistemic) justification and knowledge. The mere fact that nothing is, in principle, immune to revision should not be taken as an argument against the *a priori*.<sup>94</sup>

---

94. *A priori* truths may be necessary truths, in which case they will not turn out to be false, and will therefore not be revisable without error, either in principle or in fact. Even necessary truths are revisable in the psychological sense that we may choose to give up our belief in these truths (erroneous as this may be). Quine’s argument is presumably that even allegedly necessary truths are revisable in a stronger sense than this, i.e. that, for any statement whatsoever, no matter how necessary it may seem, there are in fact possible circumstances in which we would be *warranted* in giving it up. But even if a truth is genuinely necessary, there probably will be possible circumstances in which we would be epistemically warranted in giving it up, even if we would be in fact mistaken to do so: even if  $\langle 2 + 2 = 4 \rangle$  is a genuinely necessary truth, I might have evidence that an evil demon is merely deceiving me into accepting this statement, in which case I might be warranted in giving it up, even though I would thereby be giving up what is actually a necessary truth. Even if *a priori* truths are all necessary truths, I submit, Quine’s argument still fails.

However, there is a second Quinean argument, the Indeterminacy of Translation argument,<sup>95</sup> which may have significantly more force against the *a priori*. Quine thought, on the basis of the apparent impossibility of “radical translation” of sentences from one language into another, that psychological attitudes like belief are attitudes, not to meanings or propositions expressed by sentences, but only to the sentences themselves (79). According to Quine, “it makes no sense to speak of the translation of a single sentence of one language into another language *apart from other translations one would make* ... translation must always proceed against the background of a general scheme of translation from one language to the other” (Harman 1967, 143). As a result, we cannot say that a single sentence of one language has the exact same meaning or expresses the same proposition as a single sentence in any other language. For Quine, this indicates that meanings and propositions do not actually exist at all independently of specific sentences. And, as Bonjour explains, “[i]f sentences have no isolable meanings, if their cognitive significance is merely a function of their *de facto* connections with other sentences in the ‘web of belief’ and with experiences or stimuli,” then no sentences would have “any autonomous significance that could provide a basis for singling them out as being justified *a priori*” (1998, 79). There would be no such thing as *a priori* knowledge in this scenario, even if we recognize that *a priori* statements are (in principle) revisable in light of experience.

---

95. Bonjour thinks that this argument is better described as “the thesis of the indeterminacy of meaning and belief” (79).

BonJour, as a firm believer in *a priori* knowledge and justification, responds that this argument is simply question-begging:

the argument against the isolable meanings that *a priori* justification requires depends on assuming that knowledge is confined to empirical science, so construed as to exclude both *a priori* justification and the sort of quasi-introspective justification that would be relevant to the grasp of my own meanings. (81)

We cannot use the Indeterminacy of Translation (or the indeterminacy of meaning and belief) argument as an argument against (normative) non-naturalism if it has to assume naturalism in order to rule out the *a priori*. But why is this assumption needed as Bonjour insists that it is? The answer is that, in order to get from the observation that we cannot translate single sentences from one language into another independently of a background scheme of translation, encompassing other sentences in both languages, to the conclusion that no sentences have the autonomous meanings needed to make them knowable *a priori*, it must be assumed that users of a single sentence in one language never have anything definite in mind that might serve as the meaning of the sentence and that might be shared by users of an equivalent sentence in another language. And the only way this assumption would be plausible is if we are already entitled to make an appeal to a form of “behaviourism and verificationism, and ultimately to the ‘naturalism’ that lurks behind them” (81) according to which entities that exist inside minds, such as meanings, but which are not directly observable using the senses, therefore do not exist. But this would indeed make the argument entirely question-begging as an argument against non-naturalism and the *a priori*.

Neither of Quine's arguments could be called a knock-down argument against *a priori* knowledge in general, but the normative non-naturalist needs specifically synthetic *a priori* knowledge to avoid the second queerness argument. Perhaps the very thing that is queer about non-natural normative facts and properties is that (fundamental) knowledge of them would have to be a kind of knowledge that is not purely conceptual or linguistic in nature, but about robust features of the world, and which is nonetheless completely independent of experience.<sup>96</sup> As we observed in the previous section, it may seem that there are certain truths which are quite clearly both synthetic and *a priori*, such as *modus ponens*. Yet the naturalist can argue that while *modus ponens* is undoubtedly synthetic, it is therefore not really *a priori*, and is in fact known empirically on the basis of inference to the best explanation. Gila Sher (2013) thinks that we can differentiate between different logical systems and principles on the basis of whether or not they work in the real world (as they are used in doing things like flying airplanes, computing salaries, etc.). According to Sher, "a logical system that contains a law like *affirming the consequent* will normally not work in the world,

---

96. The idea of such knowledge makes Yudkowsky (2007b) "wonder why eyes evolved in the first place, if there are ways to produce accurate beliefs without looking at things." I think this nicely expresses the impression that there is something fundamentally "queer" about synthetic *a priori* knowledge. Of course no one is suggesting that *all* knowledge is synthetic *a priori* (even if there is such a thing as synthetic *a priori* knowledge, eyes are still an obvious epistemic necessity). But I take it that Yudkowsky's point is that *observation* and sensory organs evolved because they are necessary for substantive knowledge about the world, a fact that is arguably in tension with the idea that substantive knowledge about the world is possible without relying on any kind of observation.

while a system containing the law of *affirming the antecedent* (Modus Ponens) normally will" (160). Logical principles like *modus ponens*, for Sher, are grounded in the formal or structural properties of objects, and these formal properties have essentially the same ontological status as physical properties (172).<sup>97</sup> This suggests that the laws of logic are not actually different in kind from the laws of physics: *modus ponens* is no more of an *a priori* truth than the law of conservation of energy. And if as prime a candidate for being synthetic and *a priori* as *modus ponens* actually fails to be so, it starts to seem doubtful that there even is such a thing as synthetic *a priori* knowledge in the first place.

The non-naturalist might respond by citing apparent examples of synthetic *a priori* knowledge, such as, "no object can be red and green all over at the same time" that allegedly cannot be explained away by appeal to IBE. But if logical truths are ultimately grounded in the formal properties of objects, knowledge of which is acquired in fundamentally the same way as knowledge of physical properties, then truths like the above are probably also grounded in these same formal properties, and are thus knowable on the basis of IBE after all. Does this indicate that synthetic *a priori* knowledge should be discarded as queer, and non-natural facts and properties (normative or otherwise) with it? I don't think so. Just because certain seemingly

---

97. The formal properties of objects include such properties as the property of being self-identical, as well as cardinality and transitivity properties. Sher actually says that while physical properties hold "in limited 'regions' of the world," formal properties hold "in all its regions" (172). But in spite of this difference between the two kinds of properties, if Sher's picture is accurate, both physical and formal properties are clearly natural properties.

fundamental truths (and thus prime candidates for synthetic *a priori* knowledge) like *modus ponens* can in fact be known on the basis of IBE does not mean that they cannot also be known *a priori*. I can know that  $\langle 45, 781 + 62, 943 = 108, 724 \rangle$  either *a priori* by adding up the numbers myself, or *a posteriori* by using my calculator. In the same way, *modus ponens* may be something I can know *a posteriori* by observing how this principle works in the real world, but it still seems that I can also know that it is true *a priori* just by reflecting on and understanding the proposition that if  $\langle p \rangle$  is true, and if  $\langle \text{if } p \text{ then } q \rangle$  is also true, then  $\langle q \rangle$  must be true.<sup>98</sup>

In this context, the burden of proof is on the naturalist to show that there is something queer about synthetic *a priori* knowledge, since we are considering whether or not queerness can form the basis of a successful argument against normative non-naturalism and hence *for* normative naturalism. And even if certain truths that appear to be synthetic and *a priori* can in fact be known empirically, this does not itself establish that synthetic *a priori* knowledge is queer. Whatever reason there is to think that synthetic *a priori* knowledge is queer seems to come from the general impression that synthetic knowledge has to be knowledge of the outside world, and knowledge of the outside world has to be based on experience, and

---

98. Not to mention that an account of *modus ponens* as exclusively *a posteriori* fails to capture the apparently crucial necessary force of this “*must*.” Of course the naturalist could deny that there are any necessary truths, but once again, a necessary truth is not to be confused with a truth that is (in principle) completely immune to revision: a statement or proposition can seem to be necessary, but in fact turn out not to be in light of new evidence. This, however, does not mean that there are no necessary truths, just that we can never be absolutely certain about what the necessary truths are, just like we can never be absolutely certain about what the contingent truths are either.

therefore cannot be *a priori*. But even if this does constitute a reason to think that synthetic *a priori* knowledge is queer, it is not a strong enough reason to outweigh the reason to *believe* in synthetic *a priori* knowledge that comes from the many apparent instances of such knowledge. This includes philosophical knowledge such as (what I take to be) the very knowledge underpinning my defence of Robust Metanormative Realism in the first place: the knowledge involved in the Deliberative Indispensability Argument and Parfit's Wager. The conclusion of Parfit's Wager seems to be a synthetic *a priori* truth: something we can arrive at using a pure form of (normative) deliberation. If we accept these arguments, we should not be rejecting normative non-naturalism on the grounds that synthetic *a priori* knowledge is queer. And I don't think the arguments against synthetic *a priori* knowledge outweigh the force of these arguments for metanormative realism. So in the end, like Olson, I conclude that the second queerness argument fails.<sup>99</sup>

#### 5.4. *Queerness and irreducible normativity*

Olson himself endorses the *fourth* queerness argument, which states that (non-natural) normative facts and properties are queer because they entail that there are facts that require or favour certain responses on our part (such as adopting certain

---

99. Following his discussion of the second queerness argument, Olson moves on to the *third* queerness argument, which focuses on moral motivation, arguing that moral facts are queer because such facts would exert an inescapable motivational pull on anyone who takes themselves to be aware of them (2014, 109). I will not discuss this third queerness argument because it specifically targets morality rather than normativity as a whole (not to mention the fact that Olson rejects it anyway).

courses of behaviour), where the requiring or favouring relation is irreducibly normative, and these irreducibly normative requiring or favouring relations are queer (2014, 123-124). Irreducibly normative requiring or favouring relations “are not reducible to facts about agents’ desires, roles, or engagement in rule-governed activities” (135) and this, according to Olson, is what makes them, and consequently (non-natural) normative facts and properties, queer. Requiring or favouring relations that *are* reducible to facts about agents’ desires, roles, engagement in rule-governed activities, or the like, do not pose a similar metaphysical problem. For instance, the fact that an act is a splitting of an infinitive counts in favour of not performing that act, and this favouring relation can be reduced to the fact that not performing that act accords with a grammatical rule according to which splitting the infinitive is inappropriate, and there is not supposed to be anything metaphysically mysterious about this. Irreducibly normative favouring relations, on the other hand, “appear to be metaphysically mysterious. How,” Olson asks the non-naturalist, “can there be such relations?” (136)

If irreducibly normative favouring relations are simply reason relations that do not reduce to facts about agents’ desires, roles, or engagement in rule-governed activities, then (by the arguments of Chapters 2 and 3) we have already seen that such relations do exist, because there are such things as categorical reasons and mind-independent normative truths. As metanormative realists, we should not be rejecting these kinds of favouring relations any more than we should be rejecting synthetic *a priori* knowledge. But categoricity and mind-independence are not the

same thing as *metaphysical irreducibility*, and metaphysically irreducible normative favouring relations might be queer even if categorical normative favouring relations aren't. The Robust Metanormative Naturalist is offering a view precisely according to which there is such a thing as real or categorical normativity,<sup>100</sup> but this categorical normativity is not metaphysically primitive or *sui generis*.<sup>101</sup>

Assuming that normative facts and properties do not reduce to supernatural facts and properties, non-natural normative facts and properties clearly do entail the existence of metaphysically irreducible normative favouring relations. And such relations are metaphysically mysterious, in the sense that, being metaphysically irreducible, they do not admit of a reductive explanation that would tell us what their normativity consists in. But that does not entail that they are metaphysically *queer* in a sense that implies that such relations cannot exist; the non-naturalist could respond to Olson's question simply by asking, "Why can't there be such relations?" Olson acknowledges that the non-naturalist "could maintain that it is a fundamental

---

100. Or "objective prescriptivity," as Mackie would have said (Olson 2011, 64).

101. Although, as I will explain in Chapter 8 (Sections 8.3 and 8.4), normative properties do not actually *reduce* to non-normative properties, even though they are metaphysically dependent on them (and thus not fundamental features of reality). The non-naturalist, however, is saying something more than that normative favouring relations are irreducible either in the sense that they do not reduce to facts about agent's desires (etc.), or in the same sense in which biological properties, say, might not be completely reducible to physical properties, even though they are clearly metaphysically dependent on physical properties. The non-naturalist is saying that such relations are categorical, irreducible *and* metaphysically primitive and *sui generis*. The Robust Metanormative Naturalist believes in normative favouring relations that are both categorical and irreducible, but denies that these relations are metaphysically primitive and *sui generis*.

fact about reality that there are irreducibly normative reason relations,” and indeed this is what non-naturalists like Parfit and Scanlon actually do maintain. He also claims that, for this very reason, the issue between the naturalist and the non-naturalist “is at a bedrock metaphysical level” (136).

This seems to suggest that the Argument from Queerness simply has no force against the non-naturalist. Olson provides no argument that irreducibly normative relations are queer, beyond claiming that they appear to be so and insinuating that they require some (unforthcoming) further explanation; and he only argues that (non-natural) normative facts and properties are queer because they entail these allegedly queer relations. And the non-naturalist of course does not accept that irreducibly normative reason relations are or appear to be metaphysically queer, and nor does he accept that they are in need of any further explanation; on the contrary, such relations, for the non-naturalist, are fundamental features of reality and therefore admit of no further explanation in terms of anything more fundamental than themselves. If such relations really do exist, they probably are “mysterious,” but only in the rather tautological sense that, being fundamental, we cannot explain what they are in terms of something else. And the whole point of normative non-naturalism is that normativity cannot be reduced to or explained in terms of something else.<sup>102</sup>

---

102. Copp expresses the sentiment that “[t]heories that simply postulate sui generis normative ... properties or that help themselves to an unexplained normative notion of reasonableness or rationality are not satisfying” (2007, 2). I

In so far as there is an impasse between the naturalist and the non-naturalist (the naturalist just finds something especially queer about primitive, *sui generis* normativity, the non-naturalist just doesn't), Olson places the onus on the non-naturalist to find a way out of this impasse, perhaps by appealing to alleged companions in guilt (or innocence). The non-naturalist might say that facts about abstracta like logical and mathematical facts are no less metaphysically mysterious than (non-natural) normative facts, and since we should not deny the existence of logical and mathematical facts on grounds of queerness, we should not deny the existence of (non-natural) normative facts on queerness grounds either. And although Olson agrees that such facts about abstracta "may be metaphysically problematic in a number of ways," he responds that they do not display the feature that he finds especially queer about non-natural normative facts—they do not entail irreducibly normative reasons (138). Yet Skorupski argues that logic itself is irreducibly normative: that logical truths are actually irreducibly normative truths about epistemic reasons (2010, 203). He also argues that all *a priori* truths, including mathematical truths, are either normative truths themselves or non-normative "offshoots" of irreducibly normative truths (149).

However, as we saw in the previous section, there is a naturalistic way to account for logic in terms of formal laws which are themselves grounded in the formal properties of objects. And I think Sher's naturalistic account of logic is far

---

agree. But unfortunately, the unsatisfying nature of normative non-naturalism does not constitute a direct argument against it.

more plausible than Skorupski's normative account. Logical truths do not *prima facie* seem normative, and Sher's account has the advantage of being able to explain why logic works in the (mind-independent) world. As we saw in the previous chapter (Section 4.1, *pace* Armstrong), there is also a naturalistic way to account for mathematics in terms of the properties of classes and/or mereological wholes. If these accounts are correct, they break the analogy between abstract objects and primitive, *sui generis* normativity, because logic and mathematics would be grounded in natural properties that are knowable on the basis of observation and IBE. Unlike irreducible normativity, logic and mathematics would not be queer in the sense of being mysterious and without further explanation. But if this is the sole reason why irreducible normativity is queer and logic and mathematics are not, then to reject irreducible normativity because it is queer would be to reject it because it is irreducible, and to assume that irreducible normativity is queer (because it is irreducible) is already, in effect, to assume that non-naturalism is false. Quite aside from the success or failure of the companions in guilt strategy on the non-naturalist's part, therefore, Olson fails to make a significant case that irreducible normativity is queer, and so his fourth queerness argument in fact has no more real force than the previous three.<sup>103</sup>

---

103. As Matt Lutz and Stephen Finlay (2015) complain in their review of Olson's *Moral Error Theory*, Olson "seems to treat a response to an objection to a premise [the premise that irreducible normativity is queer and the companions in guilt response] as an argument for it. But no guilt on the charge of intolerable queerness is thereby established" (1224).

### 5.5. Irreducible normativity and the interaction problem

Nonetheless, there may be a way of explaining why non-natural normative facts and properties are queer without simply asserting, as Olson does, that irreducible normativity is queer. As I mentioned in Chapter 1 (Section 1.3), since normative properties have the capacity to exert some kind of causal influence on the natural world (so long as the argument of the previous chapter is sound), if they are not natural properties, then they must be, not just non-natural, but *supernatural*: properties that are not natural, but which are nonetheless able to affect nature, either in a brute way or by some process that is unintelligible to the natural sciences. And one can see why this might raise queerness issues: if normative properties are not a part of the natural world, how can they interact with it and influence it? Normative properties such as the property of being a reason are supposed to be able to influence our actions, and our actions, being spatio-temporal events, are not implausibly construed as simply the products of other spatio-temporal events, and it would not appear that non-natural properties can figure in any way in the spatio-temporal chains of events which typically lead to our actions. The situation for the Robust Metanormative Non-Naturalist is therefore similar to that for the mind-body dualist: if mental properties are fundamentally different from physical properties, it is difficult to see how the two can interact as they nevertheless seem to do. And if the normative is fundamentally separate from the natural, it is likewise difficult to see how the normative can influence the natural even though the normative does seem to have an influence on the natural through its influence on human action. Unless

this interaction problem admits of a solution, non-natural (supernatural) normative properties are probably best set aside as queer.

The non-naturalist might respond that the interaction problem applies no more to non-natural properties than it does to properties in general, normative or otherwise, because properties are abstract entities and so they “cannot, by definition, figure in the spatio-temporal events that make up causal regularities or chains in the way that concrete objects do” (BonJour 1998, 159). But at least some properties clearly do exert a kind of causal influence in spite of being abstract: if the fingers on my right hand did not have the property of being five, I would not see what I in fact see when I look at my right hand. The property of being five is clearly an abstract entity of some sort, but this does not prevent it from having an influence on perfectly concrete circumstances. BonJour claims that properties exert influence, not by being concretely involved in causal chains of events, but by virtue of “the presence in such a causal chain of an event or events involving concrete objects that *instantiate* these properties, where the fact that it is just those specific properties that are instantiated and not others affects the overall result” (160). And if the abstract can influence the concrete in this way, the non-naturalist could argue, there is no reason why the non-natural cannot influence the natural as well.

The objection to mind-body dualism that interaction between mental and physical properties would be impossible if these properties were fundamentally different from each other appears to be based on an appeal to at least one of three considerations: 1) a ‘billiard ball’ picture of causation, 2) the idea that the non-

physical interacting with the physical violates the principle of conservation of energy, and 3) the idea that the world is causally “closed under physics” (Robinson 2017). If the normative cannot interact with the natural because normative properties are fundamentally different from natural properties, this is probably the case for one of these three reasons. First of all, the non-natural having an influence on the natural is clearly incompatible with a “billiard ball” picture of causation, according to which causation inherently involves energy transfer through contact between concrete objects. But this picture of causation also appears to be incompatible with facts and properties having causal influence, even if the facts and properties in question are perfectly natural facts and properties. A fact is not a concrete object, so it cannot have an impact on other concrete objects the way a billiard ball can, but facts clearly seem capable of causing things: we can say truly, for instance, that the fact that the piers were broken is what caused the bridge to collapse. The “billiard ball” picture of causation, therefore, is probably invalid; it does not provide a reason why non-natural normative facts and properties might be queer.

What about the second consideration: the idea that the non-physical interacting with the physical, or the non-natural interacting with the natural, violates the principle of conservation of energy? According to Daniel Dennett (1991),

A fundamental principle of physics is that any change in the trajectory of any physical entity is an acceleration requiring the expenditure of energy, and where is this energy to come from? It is this principle of the conservation of energy that accounts for the physical impossibility of ‘perpetual motion machines,’ and the same principle is apparently violated by dualism. This confrontation between quite standard physics and dualism ... is widely regarded as the inescapable and fatal flaw of dualism. (35)

If this is an accurate assessment of mind-body dualism, the same point surely applies to Robust Metanormative Non-Naturalism as well: if normativity is not a physical or natural phenomenon, and yet it exerts a causal influence on our behaviour (which plainly involves changes in the trajectories of physical entities) the principle of the conservation of energy would seem to be violated. Can the non-naturalist simply appeal to the causal capacities of abstracta like facts and properties to get around this objection (as with the 'billiard ball' picture of causation)? If abstracta really do cause things, their doing so must be compatible with the law of conservation of energy; if it is *not* compatible with this law, we should conclude that they do not in fact cause things after all. And abstract objects are of course incapable of providing the energy needed to elicit the kinds of physical changes involved in human action.

In spite of this, natural properties clearly do figure in causal relations: dropping a square paperweight into soft clay will produce an impression, and the specific shape and depth of the impression will be caused by the shape and mass of the paperweight, shape and mass both being natural properties (Robb and Heil 2019). These properties do not expend energy in order to exert a genuine causal influence, and yet their having this influence clearly does not violate the law of conservation of energy. Physics tells us that the change in the clay must be the result of the expenditure of energy, but not that energy expenditure is the only factor involved in the specific change that takes place; properties such as shape and mass

presumably affect the *way* in which energy is expended and transferred. Human action is also the product of energy expenditure of course, and with the above in mind, we should be able to say that normative properties serve to shape and influence the way in which this energy expenditure leads to the specific actions that it does. Would this be impossible if normative properties were non-natural? Non-natural properties are properties which are both metaphysically irreducible and which can be known to be instantiated only *a priori*. And nothing about a property's being irreducible or knowable only *a priori* seems, in principle, to rule out its being able to affect nature, at least not for the reason that this would violate a law of physics such as the law of conservation of energy.

The final reason we will consider for thinking that non-natural normative properties are queer is the idea that natural (or physical) properties are the only properties able to exert any kind of causal influence, on account of the complete causal closure of physics. David Chalmers (1996) thinks that interactionist dualism about mind and body "requires a hefty bet on the future of physics, one that does not currently seem at all promising; physical events seem inexorably to be explained in terms of other physical events" (156). Robust Metanormative Non-Naturalism faces the same problem as mind-body interactionist dualism: if physical/natural causes are entirely responsible for physical/natural events, including all normatively responsive events like human actions and beliefs, then unless normative facts and properties are themselves natural, they can exert no influence at all on the natural world (and by the argument of the previous chapter, we have already seen that

normative properties do exert an influence on the world). So long as the causal closure of physics is understood in the broad sense of causality (covering all kinds of substantive influence on the world, not just spatio-temporal energy transfer or the like), it is flatly inconsistent with any form of robust non-naturalism or supernaturalism (normative or otherwise).

We should probably accept that all physical/natural events, including human actions and beliefs, have physical/natural causes. But if causal *overdetermination* is possible, this does not in fact rule out the presence of non-natural causes for physical/natural events: human behaviour, in particular, could have both physical/natural and irreducibly normative (non-natural) causes at the same time. Even though, for any human action or belief, a physical/natural event is causally sufficient to produce it, the instantiation of a non-natural, irreducibly normative property might also be causally sufficient to produce this same action or belief. For instance, I might perform a certain action *a* and be in a physical state that includes a sufficient cause of my doing *a*, such that, in a possible world in which I am not in this physical state and in which I do not do *a*, it would be true that, *had* I been in this physical state (in that world), I *would* have done *a*. But at the same time, it might also be the case that *a* has the non-natural, irreducibly normative property of being the right thing to do, and it might also be true that, in a possible world in which *a* did *not* have this property and in which I did not do *a*, nonetheless *had* *a* had this property

(in that world), I *would* have done *a*.<sup>104</sup> And so *a*'s having the non-natural normative property of being the right thing to do is in fact also causally sufficient for my doing *a*, but in a way that is compatible with every physical/natural event having a physical/natural cause.

Wedgwood (2007) employs a distinction between “world-driven” and “effect-driven” causes to defend the causal efficacy of irreducibly normative properties: “[t]he ‘effect-driven’ cause contains as little as possible that is not causally necessary in order to bring about the effect, while the ‘world-driven’ cause contains more elements and so reveals more about how the effect came about in the actual world” (195). Suppose that I have a reason to form an intention to go to London, and this normative fact obtains because of the non-normative fact that I wish to go to a concert (along with other suitable background circumstances, such as my other beliefs and desires). If I had a wish to go to a party, instead of a wish to go to a concert, I would still have a reason to form an intention to go to London, and I would still form an intention to go to London as a result. Wedgwood concludes that

the normative fact [that I have a reason to go to London] is the effect-driven cause of my forming the intention to go to London, while the purely mental

---

104. Eugene Mills (1996) discusses causal overdetermination in the context of interactionist mind-body dualism, capturing the notion of a mental event, such as a belief, being causally sufficient for a physical action, such as one's arm rising, and a physical cause *P* simultaneously being sufficient for the same action with the conditionals *O*<sub>1</sub>: “If *P* hadn't occurred but my belief had, the arm-rising would have occurred,” and *O*<sub>2</sub>: “If my belief hadn't occurred but *P* had, the arm-rising would have occurred” (107). In the normative case, the robust non-naturalist can replace the mental event—a belief—spoken of in these conditionals with the instantiation of a normative property.

[and hence non-normative] fact about the specific type and content of my mental states is the world-driven cause of my forming that intention. (196)

It would appear that physical events, such as my actually going to London having formed an intention to do so, can have both normative and non-normative causes at the same time, and thus that the presence of a physical cause for every physical event does not exclude the presence of an irreducibly normative cause for some of those physical events.

Is there anything metaphysically queer about this kind of causal overdetermination of physical events? Chalmers points out that, although such overdetermination “is often regarded with suspicion, it is hard to demonstrate conclusively that there is something wrong with it,” and that “[t]he nature of causation is sufficiently ill understood that overdetermination cannot yet be ruled out” (152).<sup>105</sup> Purely physical causal overdetermination is clearly possible: a death by firing squad is causally overdetermined—each member of the firing squad is causally sufficient to kill the target, but if you were to remove any given member, the target would still be killed anyway. Causal overdetermination as such, it would appear, is not the problem; if there is something queer about overdetermination in the case of non-natural, irreducibly normative properties, this must stem from the queerness of the properties themselves. And since the non-naturalist can appeal to overdetermination in order to get around the objection that interaction between the normative and the natural is impossible because of the casual closure of physics, in

---

105. Although he does not specifically defend overdetermination (about mental causation).

addition to the failure of the previous two reasons for thinking that normative-natural interaction is impossible under non-naturalism, the need for such interaction in fact fails to explain why non-natural normative properties are queer. What we might call the *fifth* queerness argument, therefore, based as it is on the alleged queerness of interaction between the non-natural and the natural, fails: like the previous arguments, it ends up simply assuming that non-natural normative properties are queer.

#### 5.6. Conclusion

We have looked at four different versions of the Argument from Queerness (or queerness arguments), focusing in turn on the supervenience of the normative on the natural, the (synthetic) *a priori*, irreducible normativity, and the interaction between the irreducibly normative and the natural (or between the non-natural and the natural). And we found that none of these arguments actually succeeds in explaining why non-natural normative properties and facts are (or would be) queer, at least not in any sense that entails that such facts and properties don't exist. Apart from the second queerness argument, which shoulders too heavy a burden of having to show that there is no (synthetic) *a priori* knowledge across the board, each of these arguments simply asserts or assumes that irreducible normativity is queer, and this is not enough to provide an adequate case in favour of normative naturalism. If normative scepticism was still on the table, it might be plausible to take irreducible normativity as queer, but since (I have argued) we do have good reasons to believe

in real normativity, we should not just assume that normative facts and properties cannot be fundamental, *sui generis* features of reality, given that they do exist in some form or other. Like phenomenal experience, normativity is not to be dismissed as queer just because it is difficult to explain in terms of what we already know about the world through science. To do so would be to assume, in the words of William Fitzpatrick (2016), “that we know so much about the world through scientific inquiry as to know that the default is that it should not contain value (or phenomenal experience), or that it is somehow highly unlikely that it should” (545). And I believe we do not yet know enough about the world through science to know that it is highly unlikely to contain irreducible normativity.<sup>106</sup>

But it will take a strong argument to show that normativity cannot be explained—that we have no hope of understanding it in a naturalistic manner, even if queerness does not mean that it must either be understood in such a manner or

---

106. Chalmers seems to deny the analogy between normativity and phenomenal experience when he remarks that “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” (1997, 83-84). Eliminative materialists, however, do deny that phenomenal properties exist at all. And if my argument in Chapter 3 is correct, then, so long as normative reasons are epistemically possible and we do not *know* that they don’t exist, normative reasons and normative facts do end up forcing themselves upon us. We can also observe that, as we learned from Street in Chapter 1 (Section 1.1), normative concepts are acquired from certain conscious experiences which can only be adequately described using these concepts, and so it is possible that at least some normative properties are phenomenal properties themselves. Kahane (2010) I believe correctly points out that “[t]he badness of pain seems to force itself upon us just like phenomenal properties. Indeed it imposes itself on us *through* a phenomenal property!” (47) In Chapter 7 (Section 7.7), I will argue that we cannot adequately explain our responses to certain of our conscious experience, such as agony, without appealing to normative properties.

rejected outright. The force that queerness does have against non-naturalism, I believe, reduces to the force of an appeal to Occam's Razor.<sup>107</sup> And Occam's Razor clearly provides some genuine support for naturalism: if there is no conclusive argument in favour of either naturalism or non-naturalism, naturalism will come out ahead on plausibility points precisely for reasons of parsimony (especially since it is *robust* non-naturalism we are dealing with). The Argument from Queerness itself, I conclude, is basically redundant in so far as it rests on an obvious disadvantage (robust) non-naturalistic metanormative theories have when it comes to parsimony. This disadvantage, however, might not be at all decisive, so long as there is a strong argument against all forms of normative naturalism. In the next chapter, we will see whether or not there is such an argument.

---

107. Olson himself appeals to Occam's Razor in defence of the premise that irreducible normativity is queer: if our normative "practices and beliefs can be explained without appeal to irreducibly normative properties and facts, a theory that dispenses with such properties and facts will have the advantage of being in this respect the more ontologically parsimonious theory" (2014, 147). In light of Olson's appeal to considerations of parsimony, Daan Evers (2014) is left "with the sense that queerness is not doing any work in the argument against non-natural moral [or normative] facts after all." I am left with the same exact sense.

## 6. Triviality

We have considered and rejected a potential knock-down argument against normative non-naturalism; now we will consider a potential knock-down argument against normative naturalism. Just as the Argument from Queerness is essentially based on asking: “How can there be irreducibly normative properties?” one can also ask the question: “How can natural properties be normative (or how can normative properties be natural)?” Maybe this second question can inspire a better argument than the Argument from Queerness. If normative naturalism is true, there must be some normative properties that are also natural properties (and some natural properties that are also normative properties). But many people, non-naturalists as well as error theorists and non-cognitivists, think that this is simply impossible: if a property is normative, it *cannot* therefore be natural (and *vice versa*).

Why might it be impossible for normative facts and properties to be a subset of natural facts and properties? Parfit has three interrelated arguments for this conclusion, arguments that he calls, in turn: the Normativity Objection, the Fact Stating Argument, and, most importantly (I believe), The Triviality Objection. I don’t think any of these arguments succeeds, largely for the same reason that the various versions of the Argument from Queerness don’t succeed, namely, that they all end up relying on a certain basic intuition: the queerness intuition on the one hand and

the “just-too-different intuition” on the other. And whatever force these intuitions do in fact have, I think they pretty much cancel each other out as far as the debate between the naturalist and the non-naturalist is concerned. Before I address these arguments, I will first consider the force of the non-naturalist’s *prima facie* intuition that normative properties cannot be natural properties (Section 6.1). And then I will respond to Parfit’s arguments for the fundamental impossibility of normative naturalism one at a time over the course of Sections 6.2 to 6.4.

### 6.1. *The just-too-different intuition*

Enoch (2013) insists that “[n]ormative facts and properties ... are just too different from natural ones to be a subset of them” (100). He supports this “just-too-different intuition,” as he calls it, by appealing to the perspective of the deliberating agent:

When I ask myself what I should do, it seems that just answering ‘Oh, pressing the blue button will maximize happiness’ is a complete non-starter, it completely fails to address the question. Of course, given some background commitments [such as a commitment to utilitarianism] it can be a better answer ... But such background commitments are themselves paradigmatically normative, and themselves just too different from naturalist facts and beliefs. (107-108)

But as that last sentence clearly testifies, if this line of reasoning is intended as an argument for non-naturalism, it assumes the very thing it is trying to support: that normative facts and properties are fundamentally different from natural facts and properties. After all, if some form of naturalistic utilitarianism is true, the fact that pressing the blue button will maximize happiness might actually be a normative fact

as well as a natural one. If the normative domain is autonomous from the domain of natural facts, this will of course not be the case, but that is the very thing presently at issue. To be fair, Enoch does not seem to intend the above passage as a direct argument so much as an attempt to trigger our basic intuitions; indeed he suggests that, when deciding between naturalism and non-naturalism, “[w]e may not be able to do here much more than just stare at the just-too-different intuition and try to see how plausible it seems to us, at least as a starting point” (108).

Is Enoch’s just-too-different intuition any more plausible, *prima facie*, than Olson’s intuition that irreducible normativity is queer? Perhaps it would be if normative facts and properties were completely causally inert, but not when we consider that these facts and properties do (it appears) have some kind of causal standing. In light of our ability to respond in a sensitive manner to differences in the normative facts and properties (as illustrated in Chapter 4 (Section 4.5)), I think it is in fact *prima facie* more plausible that these facts and properties are a subset of the natural facts and properties with which they seem to interact—although, as we saw in the previous chapter, interaction between natural and irreducibly normative (non-natural) properties should not be automatically ruled out as queer. Although he concedes that (robust) non-naturalism loses plausibility points due to its lack of parsimony compared to naturalism, Enoch claims that non-naturalism still comes out ahead on plausibility because naturalism conflicts with the just-too-different intuition (108-109). Yet since the normative and the natural appear to interact in some way, this intuition does not actually have the force (independently of a specific

argument for it and hence for non-naturalism itself) that Enoch takes it to have, and so, from a purely *prima facie* perspective, taking both parsimony and interaction into account, I submit that it is in fact naturalism that comes out ahead on plausibility points.

I argued in Chapter 3 that normative deliberation (like inference to the best explanation) is plausibly construed as a basic belief-forming method, meaning that we can attain at least some normative knowledge without having to rely on empirical methods. This in itself may seem to provide a knock-down argument against normative naturalism: if normative deliberation is a basic, *a priori* method of acquiring knowledge, separate from IBE or any other belief-forming method, doesn't this imply that the normative domain is separate from any other domain, including all natural and empirical domains, and thus that fundamental normative knowledge is entirely *a priori*? I don't think so. Just because normative deliberation is a basic belief-forming method, capable of yielding synthetic *a priori* knowledge (as in the case of Parfit's Wager), does not mean that fundamental normative knowledge cannot also be acquired empirically.<sup>108</sup> To pick up on the analogy with phenomenal experience introduced in the previous chapter, *introspective consciousness*—the ability to directly discern one's own conscious states—is probably a basic belief-forming method as well, but that does not mean that knowledge of conscious states cannot be acquired empirically by means of observation and IBE: my knowledge that I am

---

108. In the same way that, as we saw in the previous chapter (Section 5.3), just because a logical principle like *modus ponens* can be known empirically on the basis of IBE does not mean that it cannot also be known *a priori*.

currently thinking is not empirical knowledge, but the fact that I am thinking is still something that is knowable *a posteriori*, and through genuinely scientific methods (involving MRI machines and brain scans) at that.<sup>109</sup>

## 6.2. *The Normativity Objection*

Of the essence of the non-naturalist's position is the conviction that naturalistic accounts of normativity lose the normativity of normative facts—the very feature they are supposed to capture (Enoch 2013, 105). Olson's primary ground for rejecting normative naturalism is also that it "leaves out the irreducible normativity of normative facts" (2014, 83). Since normative facts cannot be construed in a naturalistic manner without sacrificing the normativity of these facts (for the non-naturalist), the likes of Parfit (2011b) conclude that "normative facts and natural

---

109. Some do classify introspective knowledge as empirical knowledge: BonJour (1998), for instance, claims that "[t]he justification of introspective knowledge pertaining to one's own states of mind should surely count as empirical" (7). BonJour is trying to demarcate *a priori* justification from other forms of epistemic justification, and introspection does not seem to be a source of *a priori* justification, but I do not think it should thereby be classified as *a posteriori*. Some might argue that introspection is a sense, and that introspective knowledge is therefore in the same basic category as, say, visual knowledge (i.e. both are forms of empirical knowledge). Again, I don't think this is right, because if introspection is a sense, like vision, its corresponding sense organ would presumably be the brain, and this would fail to demarcate introspective knowledge from *a priori* knowledge (one might just as well say that we have a mathematical sense, with the brain as its corresponding organ). But in any case, the point I am trying to make is that there are two ways of acquiring knowledge of conscious states, both of which involve appeals to distinct basic belief-forming methods: introspection and IBE. So just because (fundamental) normative knowledge can be acquired using an *a priori* basic belief-forming method (normative deliberation) does not mean that such knowledge cannot also be acquired using another, empirical belief-forming method (such as IBE), and thus qualify as empirical knowledge on that account.

facts are in two quite different, non-overlapping categories." If normative facts were not in a different, non-overlapping category from natural facts, the implication is, they would not, and indeed could not, be genuinely normative facts. But why should this be? Why should subsuming normative facts within the category of natural facts sacrifice the normativity of such facts? The naturalist may not be able to capture the *irreducible* normativity of normative facts, but the normative naturalist might be someone who does not believe in *irreducible* normativity at all, at least not in the metaphysical sense of "irreducible"; there would be no point objecting to such a naturalist that his view does not capture the *irreducible* normativity of normative facts.<sup>110</sup> And, according to this naturalist, normative facts do not need to be irreducibly normative in order to be normative facts, any more than biological facts must be irreducible to physical facts in order to be genuine biological facts. The naturalist Parfit is targeting is someone who thinks that, "[t]hough we make some irreducibly normative claims, there are no irreducibly normative facts" (324). So Parfit needs to show that irreducibly normative claims could not state facts that are not themselves irreducibly normative.

Parfit believes that "when we have decisive reason to act in some way, or we should or ought to act in this way, this fact could not be the same as, or consist in, some natural fact, such as some causal or psychological fact," and he defends this belief by appealing to the meanings of the words, "reason," "should," and "ought"

---

110. Though I myself will be defending a form of non-reductive naturalism (according to which normative facts are both natural and irreducibly normative) over the course of Chapters 7 and 8.

(325). But in spite of the meanings of these words, irreducibly normative claims might state natural facts in the same way that true claims about water state facts about H<sub>2</sub>O, or true claims about heat state facts about molecular kinetic energy: the meaning of “water” does not entail anything about H<sub>2</sub>O, but true claims about water nonetheless state facts about H<sub>2</sub>O. Of course it turned out that facts about water really are facts about H<sub>2</sub>O (because water is H<sub>2</sub>O), but this did not mean that facts about water are not really facts about water, nor that the *concept* of water is reducible to the *concept* of H<sub>2</sub>O. A naturalist might argue that normative facts about which actions are right are really just natural facts about which actions maximize happiness, even though the concept of being right is not reducible to the concept of maximizing happiness. In this scenario, true normative claims would be irreducibly normative (because they feature irreducibly normative concepts) and yet they would be made true by natural facts, and these natural facts, being the truthmakers for normative truths, would therefore be normative facts as well as natural facts. This suggests that normative facts can indeed be construed as natural facts without losing their normativity.

Parfit argues that normative facts cannot be like the natural facts that water is H<sub>2</sub>O or that heat is molecular kinetic energy. The concept of heat does not logically entail that heat is molecular kinetic energy, just like the concept of a right action does not logically entail that right actions are those that maximize happiness. But whereas true claims about heat are nonetheless able to state facts about molecular kinetic energy, it does not follow that true normative claims, such as claims about which

actions are right, are also able to state natural facts, such as facts about which actions maximize happiness. For Parfit, this is because the relevant concepts both leave open various possibilities and exclude various others at the same time. The concept of heat leaves open the possibility that heat could be molecular kinetic energy, but excludes the possibility that heat could be a shade of blue, or a medieval king (325). Similarly, according to Parfit, the normative concept of being a right action excludes the possibility that being a right action could be the same thing as being an action that maximizes happiness.<sup>111</sup> The point is that, even though the concept of heat does not have to entail that heat is molecular kinetic energy for claims about heat to state facts about molecular kinetic energy, these claims *can* state these facts because the concept of heat does *not* entail that heat is *not* molecular kinetic energy, although it does entail that heat is not a shade of blue. For the non-naturalist, the irreducibly normative concept of being a right action specifically excludes the possibility that being a right action is the same thing as being an action that maximizes happiness, in the same way that the concept of heat specifically excludes the possibility that heat is a shade of blue.

Of course Parfit is right that, on a conceptual level, heat could not have turned out to be a shade of blue, but that in itself does not mean that being a right action

---

111. The possibility that the concept of being a right action is supposed to exclude here is not the possibility that right actions are those that maximize happiness, in the sense that actions which have the natural property of maximizing happiness also thereby have the normative property of being a right action, but the possibility that the normative property of being a right action is the *same property* as the natural property of maximizing happiness.

could not be the same thing as being an action that maximizes happiness. The concept of being a right action no doubt excludes certain possibilities: being a right action could not turn out to be the same thing, for instance, as being an action that involves counting a certain number of blades of grass in a certain amount of time. But that does not mean that this normative property could not turn out to be the same property as *any* natural property (such as the property of maximizing happiness). Why think that the claim that being a right action is the same thing as being an action that maximizes happiness is analogous to the claim that heat is a shade of blue, rather than analogous to the claim that heat is molecular kinetic energy? Parfit says that, given the meaning of certain claims, like the claim that rivers are sonnets, or that experiences are stones, these claims could not possibly be true, and he also says that this “is the same way in which ... Normative Naturalism could not be true” (325), the implication being that, because of the meaning of the claim that being a right action is the same thing as being an action that maximizes happiness, this claim likewise could not possibly be true. But this does not actually explain why this claim, or any other fundamental normative claim of the kind that a normative naturalist might make, is analogous to obviously absurd claims like the claim that rivers are sonnets, or that experiences are stones, rather than to undoubtedly true claims like the claim that water is H<sub>2</sub>O, or that heat is molecular kinetic energy.

The Normativity Objection, it would seem, basically amounts to insisting that, because of the meaning of normative claims and the concepts involved in them,

these claims could not possibly state natural facts. Parfit considers the possibility that a normative claim, like the claim that you ought to jump from a burning hotel window into a canal in order to save your life, might state the same fact as the naturalistic claim that jumping would do most to fulfil your present fully informed desires, or is what, if you deliberated in certain naturalistically describable ways, you would choose to do. But he concludes that, given the difference between the meanings of these two claims, they could not possibly state the same fact (326), and that, on this basis alone, normative truths “could not possibly be the same as, or consist in, some merely natural fact” (327). It should be clear by now that Parfit’s Normativity Objection is just another way of stating the just-too-different intuition, and thus that it is just as inadequate as an argument against normative naturalism as a simple unvarnished appeal to this intuition (and indeed as inadequate as the Argument from Queerness is as an argument against non-naturalism). At the least, we are going to need more of an explanation as to why fundamental normative facts cannot be analogous to facts about natural kinds like the fact that water is H<sub>2</sub>O (and are instead analogous to absurd claims like the claim that heat is a shade of blue), if normative naturalism is going to be ruled out on conceptual grounds.

### 6.3. *The Fact Stating Argument*

Parfit’s second argument against naturalism might be able to explain why normative facts cannot be natural facts (like the fact that water is H<sub>2</sub>O). According to what he calls the Fact Stating Argument:

- (1) We make some irreducibly normative claims,
- (2) According to Non-Analytic Naturalists,<sup>112</sup> when such claims are true, they state facts that are both normative and natural.
- (3) If such normative facts were also natural facts, any such fact could also be stated by some other non-normative, naturalistic claim.

Therefore

- (4) Any such true normative claim would state some fact that is the same as some fact that could be stated by some other, non-normative claim.
- (5) If these two claims stated the same fact, they would give us the same information.
- (6) This non-normative claim could not state a normative fact.

Therefore

If these two claims stated the same fact, by giving us the same information, this normative claim could not state a normative fact.

Therefore

Such normative claims could not, as these Naturalists believe, state facts that are both normative and natural. (339)

If normative facts can also be stated by naturalistic, non-normative claims, as well as irreducibly normative claims, then irreducibly normative claims must give us the same information as these non-normative claims. Since non-normative claims are obviously unable to give us any normative information, if they give us the same information as irreducibly normative claims, irreducibly normative claims must not be able to give us any normative information either. Yet of course irreducibly normative claims can give us normative information. If normative naturalism entails that irreducibly normative claims state the same facts and therefore give us the same information as certain naturalistic, non-normative claims, Parfit argues, normative naturalism must be false.

---

112. Non-analytic naturalists are simply those who think that fundamental normative facts are analogous to (non-analytic) facts about natural kinds like the fact that water is H<sub>2</sub>O or that heat is molecular kinetic energy.

How might the naturalist respond to this argument? One way might be to simply deny premise (1) and maintain that even normative claims (and therefore normative concepts) are not irreducibly normative. Then, however, the naturalist would also have to maintain that fundamental normative truths are actually analytic truths that are true solely in virtue of the meanings of the terms involved. This would imply that, if a fundamental normative claim, such as the claim that right actions are those which maximize happiness, was true, this claim would be true entirely because the claim that a particular act is right means the same thing as the claim that this act would maximize happiness, in the same way that an analytic claim like “if it will rain tomorrow, then it will rain the day after today,” is true because “it will rain tomorrow” means the same thing as “it will rain the day after today.” On the most plausible account of analyticity, however, analytic claims have *no content*, that is, they do not give us any information at all, let alone normative information (Skorupski 2010, 167). This means that, if fundamental normative claims are uniformly analytic, they do not give us any normative information; and if fundamental normative claims do not give us normative information, it follows that non-fundamental/mixed normative claims (derived from fundamental normative claims plus non-normative claims) do not give us normative information either. Since, in this scenario, no normative claims would give us any normative information, such claims could not state normative facts, an outcome which is obviously fatal for the normative naturalist.<sup>113</sup>

---

113. Of course there are some analytic normative truths, such as, “a mother

Instead, the naturalist could deny premise (5) and maintain that normative and non-normative claims do not give us the same information even if they state the same facts. If we understand “fact” in the propositional sense, then of course two claims which state the same fact would give us the same information, because they would be the same proposition. But in the worldly sense, the same fact could be the truthmaker for two different claims that do not give us the same information because they present this fact in two different ways. “Hesperus is Hesperus” and “Hesperus is Phosphorus” state the same fact (in the worldly sense) but clearly do not give us the same information. In Fregean terms, these claims have the same *referent* (the fact that the planet Venus is the planet Venus, or just the planet Venus itself) but different *senses* and different cognitive value: “Hesperus is Phosphorus” allows us to get its truthmaker in mind in a way that “Hesperus is Hesperus” does not (even though they have the same truthmaker, i.e. the planet Venus<sup>114</sup>). For the naturalist, the normative claim that some act is right and the non-normative claim that this act would maximize happiness might have the same referent but different senses: they both have the same truthmaker but allow us to get this truthmaker in mind in different ways. These claims might consequently give us different

---

should do anything that all parents should do” (167). But such claims, like non-normative analytic claims like “tomorrow is the day after today,” have no informational content, and if all fundamental normative claims were like this, the entire normative domain would have no informational content.

114. Or perhaps the fact that Venus exists (in the ontological sense). See Chapter 4 (Section 4.3) for the distinction between the ontological and non-ontological senses of “exist.”

information in spite of stating the same (worldly) fact. Parfit ignores this possibility probably because, as a minimalist, he does not believe in worldly normative facts, and so he assumes that normative facts are indistinguishable from true normative propositions, and irreducibly normative propositions are obviously not equivalent to *propositional* facts which are not irreducibly normative.

But even if premise (5) is actually true, and claims which state the same fact must therefore give us the same information, I think the naturalist could still deny premise (3), and hence deny that facts that are both normative and natural must therefore be stated *either* by irreducibly normative claims *or* by naturalistic, non-normative claims. Assuming, for the sake of argument, that the way that a claim gives us a certain kind of information is by stating a certain kind of fact, if a claim states a normative fact, then it must be a normative claim, because it gives us normative information. For the same reason, if a claim states a natural fact, then it must be a natural claim. Therefore, if a claim states a fact that is both normative and natural (it gives us both normative and natural information) the claim itself must be both normative and natural. So the naturalist could say that, if a fundamental normative claim, such as the claim that right actions are those that maximize happiness, is true, then the obviously normative claim that an action is right nonetheless states a natural fact, the fact that the action maximizes happiness, and is therefore also a natural claim, and the obviously naturalistic claim that the action maximizes happiness nonetheless states a normative fact—the fact that the action is right—and is therefore also a normative claim. The first claim (that the action is

right), he might say, is *obviously* normative, but *unobviously* natural, whereas the second claim (that the action maximizes happiness) is *obviously* natural, but *unobviously* normative, since both these claims, according to the naturalist, state a fact that is both normative and natural. If the naturalist can say something like this, then Parfit's Fact Stating Argument will fail, even if we accept that different kinds of claims cannot state the same facts by presenting these facts in different ways.

Parfit might respond that the naturalist cannot actually say this, because to maintain that normative claims are also natural claims, because they state facts that are both normative and natural, would be to give up on the irreducible normativity of such claims, amounting to a rejection of premise (1) that we have already seen is untenable. If the normative claim that an action is right states the natural fact that this action maximizes happiness, this normative claim, the non-naturalist would argue, could not be an *irreducibly* normative claim. If normative claims stated natural facts, therefore, there would be no irreducibly normative claims, but, as the naturalist should agree, there clearly are such claims, and so these irreducibly normative claims could not state natural facts. Yet this response on the part of the non-naturalist simply assumes that a claim must either be natural or irreducibly normative—that it cannot be both. On the contrary, the naturalist could hold that a normative claim like the claim that an action is right *is* irreducibly normative in spite of being natural, and hence that it is both natural and, not just normative, but irreducibly normative as well: it is natural because it states a natural fact, and irreducibly normative because it uses an irreducibly normative concept (the concept

of being right). For the naturalist, a claim can be both natural and irreducibly normative precisely because the normative and the natural do not constitute fundamentally separate categories or domains.

The non-naturalist might still object that, if an irreducibly normative claim, like the claim that an action is right, states a natural fact, like the fact that this action would maximize happiness, then, since this fact can also be stated simply by the claim that this action would maximize happiness, *this* claim must give us the same information as the irreducibly normative claim that the action is right. But if it did, then the claim that the action would maximize happiness would not only be a normative claim, but an *irreducibly* normative claim. And if this claim is an irreducibly normative claim, just like the claim that the action is right, then it must contain at least one irreducibly normative concept. Yet a clearly naturalistic claim such as the claim that an action maximizes happiness does not feature any irreducibly normative concepts. Therefore, according to the non-naturalist, this claim cannot be an irreducibly normative claim; hence it cannot convey the same information, or state the same fact, as the irreducibly normative claim that the action is right.

If the claim that an action maximizes happiness is a normative claim as well as a naturalistic one, then it must contain at least one normative concept: if a claim is both natural and normative, it must contain at least one concept that is natural and at least one concept that is normative. And this should not pose a problem for the naturalist so long as there is already no problem with saying that some obviously

naturalistic claims are nonetheless (unobviously) normative—the concept of happiness, for instance, while clearly a naturalistic concept, could also be a normative concept (albeit unobviously). But could any such natural concept also be an *irreducibly* normative concept? A concept like happiness, it would seem, may be both natural and normative, but if it is indeed natural, it might be difficult to see how it could also be *irreducibly* normative. And if it could not be, the claim that an action maximizes happiness could not be an irreducibly normative claim, in which case it could not give us irreducibly normative information, and this threatens to entail the conclusion that it could not state the same fact as an irreducibly normative claim (such as the claim that the action is right).

I don't think this conclusion actually follows, however. On the one hand, it could be that all normative concepts are irreducibly normative, even if normative facts and properties are not, in which case, if the concept of maximizing happiness is indeed a normative concept, it is also an irreducibly normative one, just like the concept of being right. Perhaps the concept of maximizing happiness cannot be adequately explained or analyzed without invoking (other) normative concepts. But even if naturalistic concepts are not irreducibly normative, for the naturalist that Parfit is arguing against, the non-analytic naturalist, there are no irreducibly normative facts, so even irreducibly normative claims do not give us *irreducibly* normative information in the sense of stating irreducibly normative facts. For such a naturalist, the claim that an action maximizes happiness does not need to give us irreducibly normative information (in the sense of stating an irreducibly normative

fact) in order to state the same fact as an irreducibly normative claim (such as the claim that the action is right). So the issue becomes whether or not a claim that *is* irreducibly normative can state a fact which is not irreducibly normative. According to the non-analytic naturalist, the situation precisely is that facts that are not irreducibly normative can be and often are stated by irreducibly normative claims, because it is these claims, and the concepts featured in them, which are irreducibly normative, and not the (natural and normative) facts which they state or the (natural and normative) properties involved in these facts. This cannot be assumed to be impossible without begging the question against the non-analytic naturalist. And so I conclude that, like the Normativity Objection, the Fact Stating Argument fails.

#### 6.4. *The Triviality Objection*

Parfit's third and final argument against normative naturalism, the Triviality Objection, states that, if naturalism was true, and normative facts and properties were also natural facts and properties, no normative claim that is both *substantive* and *positive* could be true. In Parfit's terms, normative claims are substantive and positive if they "state or imply that, when something has certain natural properties, this thing has some other, different, normative property," and if they "are significant, because we might disagree with them, or they might tell us something we didn't already know" (2011b, 343). And if there were, in this sense, no substantive, positive normative truths—no truths which tell us that something actually has a certain normative property, and which, in doing so, also tell us

something significant that we might not already know — there would not be any irreducibly normative truths either, from which it follows that no irreducibly normative claim could state any fact at all, let alone a natural fact (or a fact which is both natural and normative). This means that, if normative facts and properties are natural facts and properties, irreducibly normative claims could not state natural facts, but normative naturalism requires both that normative facts and properties are natural and that irreducibly normative claims can state such facts. Normative naturalism, therefore, must be false (if, that is, Parfit is right that no substantive, positive normative claim could be true if normative facts and properties were also natural facts and properties).

Why couldn't there be any substantive, positive normative truths (in Parfit's sense) if normative facts and properties were also natural facts and properties? Parfit takes two paradigmatic claims that a normative naturalist might make: the (apparently) first-order normative claim (A) that "when some act would maximize happiness, this act is what we ought to do," and the second-order, metanormative claim (C) that "when some act would maximize happiness, this property of this act is the same as the property of being what we ought to do" (343). He then argues that:

- (1) (A) is a substantive normative claim, which would, if it were true, state a positive substantive normative fact.
- (2) If ... (C) were true, (A) could not state such a fact. (A) could not be used to imply that, when some act would maximize happiness, this act would have the different property of being what we ought to do, since (C) claims there is no such different property. Though (A) and (C) have different meanings, (A) would only be another way of stating the trivial fact that, when some act would maximize happiness, this act would maximize happiness. (343-344)

Premises (1) and (2) form an inconsistent triad with the naturalist's claim (C): by premise (2), if (C) was true, (A) would not be a substantive, positive normative claim (in Parfit's sense)—it would not tell us that acts which maximize happiness also have the *different* property of being what we ought to do (because, according to (C), maximizing happiness is the same property as being what we ought to do). Yet, by premise (1), (A) is a substantive, positive normative claim. Therefore, either claim (C) is false, along with any equivalent metanormative claim that a naturalist might make, meaning that no normative property is the same as any natural property, or premise (1) is false, and claim (A), along with any other naturalistic, (apparently) first-order normative claim of this kind, is *not* in fact a substantive, positive normative claim. So if there were no non-natural, irreducibly normative properties, and all normative properties were the same as certain natural properties, there would be no substantive, positive normative truths, and therefore no irreducibly normative truths.

Jackson's response to this argument is to defend the significance of identity claims between natural properties and normative properties. According to Jackson (2016), to identify a certain natural property with the normative property of being right is not just to make the trivial claim that some property is the same as itself, but rather to say that this natural property fulfils a certain job description specified by the concept of being right, and since "it takes real work to find the natural property to identify with being right ... the identification isn't trivial" (207). He gives the

analogy of the concept of a random sequence in probability theory and statistics, where what it takes to be a random sequence is a controversial and difficult question. In this case, “when someone offers an account of what it takes to be a random sequence, the proposal isn’t that there are two properties, that of being a random sequence and that of being so and so, where so and so is the account on offer ... [Rather, the proposal is that] there’s one property” (208). And yet such an account, whatever it might be, will clearly be far from trivial. In the normative case, therefore, the presence of only one property, to which concepts such as maximizing happiness and being right both refer, will not make the resulting identity claim trivial either.

Identity claims between natural and normative properties are not trivial claims, even if they do effectively say that some property is the same as itself. Claim (A), no doubt, is not a trivial claim, even if claim (C) is true. If (C) is true, then (A) would not so much be another way of stating the trivial fact that, if some act would maximize happiness, this act would maximize happiness, but rather another way of stating the fact that (C)—that maximizing happiness is the same property as being what we ought to do. And (C) is clearly not a trivial claim: we might not know that being what we ought to do is the same property as maximizing happiness (indeed Parfit doesn’t know this) and this is something it might be hard for us to learn. But just to point out that either claim (A) or claim (C) is not a trivial claim is not an adequate response to Parfit’s argument, because although the argument is called the Triviality Objection, the objection it actually makes to normative naturalism is not

that it would make claims like (A) trivial, but that it would prevent such claims from being substantive, positive normative claims, and that it therefore implies that, not only are there no irreducibly normative facts and properties, but no irreducibly normative truths either. And, according to premise (1) of Parfit's argument, claim (A) is a substantive, positive normative claim, and as we have already seen, even if there are no irreducibly normative facts (in the worldly sense), there must be some irreducibly normative truths if the normative domain is to have any meaningful content at all.

If it is not enough to simply defend the non-triviality of identity claims like (C), how might the naturalist respond to the so-called Triviality Objection? The first thing to note is that whatever force this argument does have applies only to *reductive* forms of naturalism: those naturalistic theories which purport to metaphysically reduce normative properties to natural properties by identifying certain uncontroversially normative properties (such as the property of being what we ought to do) with certain uncontroversially natural properties (such as the property of maximizing happiness). It does not address non-reductive forms of naturalism which do not rely on identity claims between normative and natural properties like (C). Even if no claim like (C) was true, and no (uncontroversially) normative property was the same as any (uncontroversially) natural property, normative naturalism could still be true in the epistemic sense that fundamental normative knowledge can be acquired empirically and not just *a priori* as the non-naturalist believes. If we could know on the basis of inference to the best explanation that right

actions are those which maximize happiness, normative naturalism would still be true even if the property of being right was a completely different property from the property of maximizing happiness, and indeed was a different property from any uncontroversially natural property. As we observed in Chapter 1 (Section 1.3), Epistemic Normative Naturalism is not the same thing as (metaphysical) normative reductionism.

But even as an argument against reductionism, the Triviality Objection, I believe, still ultimately fails: even if a metanormative claim like (C) is true, this would not mean that there are no substantive, positive normative truths, nor that there are no irreducibly normative truths. Remember, according to Parfit, a normative claim is both substantive and positive when it tells us that something with certain natural properties has some other, different, normative property.<sup>115</sup> And if (C) is true, and the property of being what we ought to do is the same property as the property of maximizing happiness, then the property of maximizing happiness will itself be a normative property as well as a natural property, just like the property of being what we ought to do. So if a truth tells us that some act, which has certain (other) natural properties, also has the normative (and natural) property of

---

115. "Different" here cannot mean different from all natural properties if Parfit's criteria for a substantive, positive normative truth is not to be question-begging. It must therefore simply mean different from certain other natural properties that the thing in question has; if a normative claim says that an act with a certain natural property also has a certain normative property which is different from that natural property, even if this normative property is itself a natural property, then this normative claim will come out as substantive and positive on Parfit's definition.

maximizing happiness, then this will be a substantive, positive normative truth in Parfit's sense. And it will also be a fundamental normative truth, since it will tell us that, whenever something has certain non-normative properties, such as the kind of non-normative properties that make an act have the property of maximizing happiness, it will thereby also have a certain normative property, such as (perhaps) the property of maximizing happiness.<sup>116</sup>

There are certain identity claims between properties, at least one of which is a normative property, that do not seem to raise a problem for Parfit's own version of non-naturalism. After all, Parfit himself (2011a) implies that the normative property of *being a reason for* something is the same property as the property of *counting in favour of* that thing (31).<sup>117</sup> While this does make certain normative claims at least somewhat trivial (for instance, the claim that, if something is a reason for an act, then that thing also counts in favour of that act), it is far from making all fundamental normative claims trivial—any claim that says that something is a reason for an act, or that something counts in favour of an act, will be a substantive, positive normative claim, because both these kinds of claims informatively state that something (which no doubt has various non-normative properties) also has a particular normative

---

116. Recall Streumer's definition of fundamental normative truths (2017) as "normative truths that obtain independently of which objects have which descriptive [or non-normative] properties" (34).

117. In his words, "[i]t 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 acting in some way. But 'counts in favour of' means roughly 'gives a reason for'" (31).

property, even if this property is in fact the same property as some “other” (normative) property. And this is exactly the scenario that obtains under (reductive) normative naturalism: fundamental normative claims say that something with various non-normative properties also has a particular normative property (such as the property of being what we ought to do), which may be the same as some “other” normative property (such as, for some naturalists, the property of maximizing happiness), even if this property is also a natural property. Even for Parfit, the truth of “normative property = normative property” identity claims does not threaten the substantive and positive status of fundamental normative truths (nor indeed do “natural property = natural property” identity claims); it is supposed to be only “normative property = natural property” identity claims that do this. But, for the reductive naturalist, “normative property = natural property” identity claims *are* “normative property = normative property identity claims” (and “natural property = natural property” identity claims as well, of course). And since these identity claims do not undermine the substantive and positive status of fundamental normative claims, naturalistic metanormative claims like (C) need not do so either, and thus the truth of a claim like (C) should not imply that there are no irreducibly normative truths.

Parfit might still object that, even though reductionism does in fact allow for substantive positive normative claims and irreducibly normative truths, it would still imply that claim (A) is not a substantive, positive normative claim, and since (A) *is* a substantive, positive normative claim (by Parfit’s premise (1)), reductionism

must be false. After all, the claim that, when some act would maximize happiness, this act is what we ought to do, certainly looks like a first-order normative claim that is both substantive and positive. But if reductionism is true, and hence claim (C) is true, claim (A) would be no more of a substantive, positive normative claim than the claim that, when something is a reason for an act, that thing counts in favour of that act, or the claim that, when some act is what we ought to do, this act is right.<sup>118</sup> And these claims are not substantive, positive normative claims, and so, contrary to appearances, (A) would not be a substantive, positive normative claim either.

Embracing reductive naturalism would therefore mean giving up on the substantive, positive normative status of certain claims like (A) which nonetheless seem to possess this status. For this reason, reductionism does involve some loss of plausibility points—having the somewhat implausible implication that the utilitarian claim that, when some act would maximize happiness, this act is what we ought to do, does not actually give us any positive normative information—but this is far from a knock-down argument against reductionism, and we have already seen that non-naturalism comes with its own loss of plausibility points in other areas, parsimony being foremost among them.

---

118. Claim (A) may be more substantive than these claims in the sense that it is less obviously true than they are, but it would not be any more of a positive normative claim, because none of these claims tells us that something has some normative property because it has some different, non-normative property.

### 6.5. Conclusion

The Triviality Objection fails as an argument against normative naturalism for basically the same reason that the Normativity Objection and the Fact Stating Argument fail: they all end up simply assuming that natural facts and properties cannot be normative facts and properties. If an uncontroversially natural property, like the property of maximizing happiness, was also a normative property (albeit unobviously), then a property identity claim such as “maximizing happiness is the same thing as being what we ought to do” could well be analogous to a natural kind identity claim like “water is H<sub>2</sub>O,” and not to an obviously absurd claim like “heat is a shade of blue.” If the natural fact that some act would maximize happiness was also a normative fact, this fact could well be stated by an irreducibly normative claim, such as the claim that this act is right. If maximizing happiness was also a normative property, claims that say that some act maximizes happiness would be substantive, positive normative claims, because they would say that something, which no doubt has certain non-normative properties (or *other* natural properties), also has a certain normative property (even if this property is also a natural property), just like the claim that some act is what we ought to do. And if the claim that some act maximizes happiness could indeed be a substantive, positive normative claim, then the claim that this act is what we ought to do could be an irreducibly normative truth, even if what makes it true is an entirely natural fact: the fact that this act would maximize happiness.

Of course one might think that a natural property like maximizing happiness is just too different from any normative property to be a normative property itself, but in light of the fact that normative properties very much seem to interact with natural properties, this just-too-different intuition just does not have enough force to provide good support for non-naturalism. One might also wonder what it would even mean for a property like maximizing happiness to be a normative property, but if maximizing happiness really is a normative property, this could be explained precisely by the fact that this property is the same property as some obviously normative property, such as the property of being what we ought to do. Thus I conclude that the kind of conceptual and metaphysical considerations that Parfit appeals to in his arguments are not in fact a problem for normative naturalism. If non-naturalism is actually the more plausible view, I believe, this must be for epistemological, rather than metaphysical, reasons. In the next chapter, therefore, we will examine the epistemological case for and against normative naturalism.

## 7. Normative Explanations

If normative facts and properties can interact with and affect non-normative facts and properties, normative facts should also be able to *explain* why certain non-normative facts obtain. If a certain non-normative fact is in part counterfactually dependent on a certain normative fact,<sup>119</sup> then part of what explains why that particular non-normative fact obtains is not just other non-normative facts but also a normative fact. As such, we should sometimes have reason to believe that a certain normative fact obtains because it is an indispensable part of the best explanation for a certain non-normative fact (or facts). And if normative knowledge can be acquired in this way through inference to the best explanation from non-normative, empirical facts, this knowledge will itself qualify as a branch of empirical knowledge, not different in kind from knowledge of the “irreducible posits” of natural science, like subatomic particles and forces of nature (Quine 1980, 44-45).

We have seen that many of our responses do seem to be counterfactually dependent on normative facts and properties (Chapter 4, Section 4.5). But nonetheless, many think that the empirical, explanatory model of knowledge appropriate to the natural sciences is simply inapplicable to the normative domain.

---

119. If I would not have stopped my car at the red light “but for” the normative fact that I had a reason to stop at the red light, to return to Temkin’s example (2016, 5).

And this, I believe, is the main challenge for normative naturalism: to explain how we can draw plausible normative conclusions from non-normative, empirical facts alone, without relying on implicit, *a priori* normative assumptions in order to do so. Hume thought that no normative conclusions could be drawn from purely non-normative, empirical premises without relying, implicitly or explicitly, on at least one purely normative, non-empirical premise—you cannot go from an “is” to an “ought,” as the saying goes. If I ought to push the blue button, Hume would say, this cannot be concluded just from the natural fact that pushing the blue button would maximize happiness; we also need the purely normative premise that what I ought to do is maximize happiness. And how can we conclude that what I ought to do is maximize happiness (or anything else for that matter) on the basis of empirical evidence alone, without relying on some further normative premise or taking that normative premise itself as epistemically basic, and hence as *a priori* and non-natural?

In my attempt to answer this question and meet what we might call the “Humean challenge” to normative naturalism, I will start by considering Sturgeon’s moral explanations argument in Section 7.1. Although I of course agree with Sturgeon that there are correct normative explanations for empirical facts, Sturgeon’s well-known examples do not actually show what the naturalist needs to show: that *fundamental* normative knowledge can be acquired on the basis of observation and IBE. For the non-naturalist, fundamental normative facts play no role whatsoever in explaining non-normative, empirical facts, and thus IBE is never the method we use

to acquire fundamental normative knowledge. For some non-naturalists, normative knowledge is always the product of the “reflective equilibrium” method originally described by John Rawls. In Scanlon’s words (2003), “this method, properly understood, is ... the best way of making up one’s mind about moral [or normative] matters ... Indeed, it is the only defensible method: apparent alternatives to it are illusory” (149). Other non-naturalists endorse an intuitionist model of normative knowledge, according to which certain fundamental normative truths are epistemically basic and self-evident, and may even be conceptual truths. According to Cuneo and Shafer-Landau (2014), certain moral propositions, which they call “the moral fixed points,” are conceptual truths: truths that are known *a priori* solely on the basis of understanding the concepts involved (normative and non-normative) and the necessary relations between them (405).<sup>120</sup> We do not arrive at knowledge of these truths either by using IBE or through a process of seeking reflective equilibrium; rather, we immediately and intuitively grasp that the concepts involved in these propositions necessitate that they must be true.

In Sections 7.2 and 7.3, I will argue that the reflective equilibrium method cannot be taken as a basic method of forming justified normative beliefs, and that, as far as the non-naturalist is concerned, it must be underpinned by some form of foundationalist intuitionism. In Section 7.4, I will argue that the non-naturalist should regard fundamental normative truths as conceptual truths, after the fashion

---

120. A couple of examples of these so-called moral fixed points are: “It is pro tanto wrong to engage in recreational slaughter of a fellow person,” and “It is pro tanto wrong to torture others just because they have inconvenienced you” (405).

of Cuneo and Shafer-Landau, but that the substantive nature of the conceptual truths in question means that the non-naturalist still needs a further account of how these truths are known in the first place. To this effect, in Section 7.5, I will outline what I believe to be the most plausible non-naturalistic account of normative knowledge: Skorupski's *spontaneity and convergence* account, according to which warranted normative beliefs are those that are based on normative dispositions which are not the product of "debunking" causal explanations and which converge sufficiently well with the normative dispositions of others. I will then supply a critique of this account in Section 7.6, arguing that, on its own terms, it actually ends up implying that normative knowledge *can* be acquired on the basis of IBE, and thus that there are correct normative explanations for non-normative, empirical facts. This in turn implies that our normative dispositions are themselves causal responses, contrary to the very hypothesis of normative non-naturalism.

Finally, in Section 7.7, I will explicitly make the argument that IBE can take us to knowledge of fundamental normative truths. In particular, I will argue that we cannot adequately explain why conscious beings respond in the specific ways that they do to the specific kinds of feelings and sensations they experience without positing that these feelings and sensations have normative properties. We can know the fundamental normative truth that agony always gives us a reason to want and to try to avoid it, because this offers the most plausible explanation of why conscious beings (at least those that actually know what agony is like) universally want to avoid agony.

The thrust of my argument is partly inspired by Mill's famous (or infamous) "proof" of the Principle of Utility. As a proponent of the "inductive" school of ethics, according to which "right and wrong, as well as truth and falsehood, are questions of observation and experience" (50), Mill (1998) claimed "that happiness is desirable, and the only thing desirable, as an end." In defence of this principle, he argued that

The only proof capable of being given that an object is visible, is that people actually see it. The only proof that a sound is audible, is that people hear it: and so of the other sources of our experiences. In like manner, I apprehend, the sole evidence it is possible to produce that anything is desirable, is that people actually desire it. (81)

In Humean fashion, Russell dismissed this reasoning as committing an obvious fallacy. As Russell (2004) understands Mill, "[h]e does not notice that a thing is 'visible' if it can be seen, but 'desirable' if it *ought* to be desired. Thus 'desirable' is a word presupposing an ethical theory; we cannot infer what is desirable from what is desired" (741).

What Russell did not notice, however, is that Mill was not making a *deductive* inference from the premise that happiness is desired to the conclusion that it is desirable: of course the (empirical) fact that happiness is desired does not *entail* the (normative) conclusion that happiness is desirable. But Mill is more charitably (and more plausibly) interpreted as meaning that this fact constitutes *pro tanto evidence* for the conclusion that happiness is desirable, not that it deductively entails it. In his words:

No reason can be given why the general happiness is desirable, except that each person, so far as he believes it to be attainable, desires his own happiness. This, however, being a fact, we have not only all the proof which

the case admits of, but all which it is possible to require, that happiness is a good. (1998, 81)

And I think Mill was right: the fact that happiness is desired does constitute good evidence for the conclusion that happiness is desirable. If happiness ought to be desired as an end in itself, it must be the very nature of happiness that makes it the case that happiness ought to be desired in this way. And if happiness ought to be desired, not as a means to something else but for itself by its very nature, we would expect the kinds of beings who (to some extent at least) are capable of experiencing happiness, who know by direct experience what happiness is like, and who are therefore well acquainted with the intrinsic nature of happiness, precisely to have a strong desire for happiness. If it turned out, on the contrary, that these beings did *not* have any desire for happiness at all, this would surely falsify the hypothesis that happiness is intrinsically desirable: it would be incredible that beings who know exactly what happiness is like (and who are not completely insane, let's say) nonetheless fail to have any desire for happiness if happiness is indeed intrinsically desirable. But if, on the other hand, these beings *do* universally possess a strong desire for happiness—as a result of their direct acquaintance with happiness and its nature—I believe this is evidence for the hypothesis that happiness actually is desirable: it confirms a falsifiable prediction that this hypothesis genuinely seems to make. And if there is no more plausible alternative explanation for the fact that happiness is desired (and hence no significant evidence *against* the hypothesis that

happiness is desirable), this evidence should be enough to indicate that this hypothesis is likely to be true.

Yet I think we can ultimately vindicate normative naturalism without relying on Mill's contentious assumption that happiness—defined simply as “pleasure, and the absence of pain” (55)—is the only thing desired for its own sake (rather than, say, freedom, integrity, or meaning). Even if it is far from clear that everyone desires a positive balance of pleasure over pain, it seems perfectly clear that everyone has some desire to avoid intense agony. And if agony is intrinsically undesirable and by its nature ought to be avoided for its own sake, this is exactly what we would expect to find: that beings like us, who know exactly what it's like to experience agony and who are thus directly acquainted with its nature, would always have at least some desire to avoid agony. This observation should therefore give us hope of placing normative knowledge on a firmly empirical basis.

### *7.1. Normative facts and best explanations*

Taking after the “inductive” tradition of Mill, Nicholas Sturgeon (1988) purports to show that there can be empirical evidence for normative facts on the grounds that normative facts feature in the best explanations for certain non-normative, empirical facts. According to Sturgeon, the best explanation (or part of the best explanation) for the non-normative fact that

vigorous and widespread moral opposition to slavery arose for the first time in the eighteenth and nineteenth centuries, even though slavery was a very old institution ... [and that] this opposition arose primarily in Britain, France,

and in French- and English-speaking North America, even though slavery existed throughout the New World ... is [the normative fact] that chattel slavery in British and French America, and then in the United States, was much *worse* than previous forms of slavery, and much worse than slavery in Latin America. (245)

If chattel slavery in North America in the eighteenth and nineteenth centuries had not been much worse than previous forms of slavery, there would not have been such vigorous and widespread opposition to slavery at that time. The severe badness of this form of slavery was presumably not the only factor that contributed to such vigorous and widespread opposition. But if there really are normative facts, and if these normative facts really do have causal standing (as I argued in Chapters 2, 3 and 4), then surely this normative fact was at least one of the contributing factors to this opposition. And since we know on empirical grounds that opposition to slavery was indeed as vigorous and widespread as it actually was (at that time), we are also in a position to know the normative fact that chattel slavery in North America was much worse than previous forms of slavery on the basis of observation and IBE, without having to rely on *a priori* normative assumptions.<sup>121 122</sup>

---

121. One could argue that inference to the best explanation is itself an *a priori* normative truth, but as we suggested in Chapter 3, the reason to accept IBE as a legitimate or reliable form of inference in the first place probably derives from a combination of empirical observation and IBE itself (the observation that IBE usually works by leading us to have true beliefs, plus inference to the best explanation of *that* observation that IBE is probably a legitimate or reliable form of inference), and thus it is not a purely *a priori* truth, even if it is a normative one.

122. Those who actually engaged in the vigorous and widespread opposition to slavery at that time presumably were not basing their knowledge that slavery is wrong on IBE from the empirical fact of their own opposition, but the point of the example is that normative truths *can* be known on the basis of IBE, not that they

Examples like this seem clearly to show that normative knowledge can be acquired *a posteriori*. But even if we do know certain normative facts, such as the fact that chattel slavery in eighteenth and nineteenth century North America was particularly bad, or that Hitler was a morally depraved person (234), on the basis of IBE,<sup>123</sup> this will not be enough to establish that normative knowledge is in-and-of-itself a form of empirical knowledge. As in the debate between the normative minimalist and the Robust Realist, it is the status of *fundamental* or pure normative truths and facts that is at issue between the naturalist and the non-naturalist.<sup>124</sup> A normative fact like the fact that Hitler was morally depraved is not a fundamental but rather a mixed normative fact, and mixed normative facts can of course be known empirically because they are in part empirical facts: if we didn't have empirical knowledge about who Hitler was and what he did, we wouldn't know that Hitler was a morally depraved person, so this normative fact cannot be known purely *a priori*. But just because mixed normative facts, which partly depend on uncontroversially empirical facts, are not purely *a priori* does not mean that

---

must be. And even for those who opposed slavery at the time, one of the things that caused them to believe that slavery was wrong would have been the normative fact that slavery is wrong, so their knowledge would have had a causal basis even if it was not based on IBE.

123. In the Hitler case, from the empirical fact that he did the things that he did.

124. The epistemic status in the present case, the ontological status in the previous case.

fundamental normative facts are not *a priori* or that they can ever be known empirically.

## 7.2. *Reflective equilibrium*

Let's turn to the reflective equilibrium method. What exactly does this method involve? For the non-naturalist, remember, the normative is a completely autonomous domain; Gilbert Harman (2000) describes the approach involved in what he calls "autonomous ethics" as one in which

We begin with our initial moral beliefs and search for general principles. Our initial opinions can be changed to some extent so as to come into agreement with appealing general principles and our beliefs about the facts, but an important aspect of the appeal of such principles will be the way in which they account for what we already accept. (81)

For Scanlon (2014), this basic method applies not only to the moral or normative domains, but to non-normative domains like set theory as well. As he describes it, "[o]ne begins by identifying a set of considered judgements, at any level of generality, about the subject in question. These are judgements that seem clearly to be correct and seem so under conditions that are conducive to making good judgments of the relevant kind about this subject matter." Next, "we formulate general principles that would 'account for' these judgements ... principles such that, had one simply been trying to apply them, rather than trying directly to decide what is the case about the subject at hand, one would have been led to the same set of judgements." If these general principles conflict with our (specific) judgements, we must decide "whether to give up the judgements that the principles fail to account

for, to modify the principles, in hopes of achieving a better fit, or to do some combination of these things" (77). And we continue this process indefinitely until we achieve a state of coherence between our general principles and our specific judgements, at which point we are justified in concluding that our normative principles and judgements constitute knowledge.

The reflective equilibrium method has coherentist implications for normative knowledge. In this process, although we may start with certain specific judgements or general principles that are not inferred from anything else, these judgements and principles can always be given up if they turn out to conflict with other judgements or principles that we find more plausible. No normative judgement or principle, this suggests, is truly foundational or self-evident. The non-naturalist who defends the reflective equilibrium method may argue that it is enough if our normative judgements and principles are consistent, with each other and with our (justified) non-normative beliefs, for them to qualify as knowledge, although none of them can be considered foundational or self-evident. Even if there are no self-evident normative truths, there may be no self-evident non-normative truths either, in which case, our normative beliefs would be epistemically on par with our non-normative beliefs, so long as these two sets of beliefs are internally consistent and do not conflict with each other. If coherentism is broadly the correct theory of knowledge, and thus there is no need for our scientific beliefs to possess some kind of ultimate justification prior to the natural sciences themselves, there may also be no need for our normative beliefs to be justified in non-normative terms (by being supported by

or inferred from purely empirical premises, for instance, as the naturalist thinks they can be). And if scientific knowledge does not rest on a bedrock of self-evident non-normative, empirical truths, normative knowledge might not require a bedrock of self-evident normative truths either.<sup>125</sup>

Yet the non-naturalist requires not only that normative knowledge can be acquired using the reflective equilibrium method, but also, as Scanlon says, that there is no alternative to this method—that fundamental normative knowledge cannot also be acquired using IBE.<sup>126</sup> Even if our normative beliefs are justified so long as they cohere with each other and do not conflict with our (justified) non-normative beliefs, this does not mean that fundamental normative knowledge cannot be acquired empirically. It is not incompatible with normative naturalism that some normative knowledge can be acquired just by reflection, without relying on empirical observation or the explanation thereof. Normative truths that are knowable *a priori* might also be knowable through naturalistic means, or there may be other (fundamental) normative truths that can be known empirically, even if there are some purely *a priori* normative truths that can't.

---

125. Such as Cuneo and Shafer-Landau's moral fixed points.

126. That is, knowledge of fundamental or pure normative truths. The non-naturalist should admit that knowledge of mixed normative truths can be acquired using IBE, specifically in order to determine the specific (non-normative) facts of the case on which these mixed normative truths depend.

### 7.3. *The need for intuition*

For the non-naturalist, of course, fundamental normative truths are known *a priori*, and only *a priori*. But not all normative judgements can be known to be true *a priori*: when two people make conflicting normative judgements, at most one of these judgements is correct (Streumer 2017, 6-7).<sup>127</sup> How do we tell which (fundamental) normative judgements represent genuine *a priori* truths, or which we are warranted in taking to be genuine *a priori* truths? Parfit (2011a) takes it to be an *a priori* truth that, in his words, “[w]e all have a reason to want to avoid, and to try to avoid, all future agony” (76), even in cases where “I know that some future event would cause me to have some period of agony,” and “[e]ven after ideal deliberation, I have no desire to avoid this agony. Nor do I have any other desire or aim whose fulfilment would be prevented either by this agony, or by my having no desire to avoid this agony” (73-74).<sup>128</sup> When he considers the possibility that someone might deny this and instead claim that, in cases like the above, “though the approaching flames threaten to cause me excruciating pain, this fact does not count in favour of my wanting and trying to move my hand away,” his response is simply to say: “that is hard to believe” (81).

---

127. Streumer takes this claim to be one of the *central thoughts* about normative judgements and properties, which reflects the nature of these judgements and properties (6-7).

128. The point being that we have a reason to want to avoid, and to try to avoid, agony, not just because avoiding agony would fulfil our desires, but because of the nature of agony itself.

It is not enough simply to appeal to reflective equilibrium here. As Scanlon himself says, “the justificatory force of the fact that we have arrived at certain judgements in reflective equilibrium depends on the substantive merits of the judgements we make along the way” (2014, 82). After all, “different people, applying this method about the same subject, may reach different reflective equilibria” (79), and thus accept conflicting normative judgements which, by this criterion alone, would both be equal candidates for *a priori* knowledge. Parfit would probably say that his normative judgement that we all have a reason to want to avoid, and to try to avoid, future agony (let’s call it “the Agony Principle”) has the same basic epistemic status as logical truths like *modus ponens* or the law of non-contradiction, and thus it needs no further justification or evidence to support it.<sup>129</sup> But as we learned from Sher (2013) in Chapter 5 (Section 5.3), logical truths may not be epistemically basic and *a priori*, but known empirically on the basis of IBE (from how logical principles work in the real world as they they are used in doing things like flying airplanes, computing salaries, etc.) (160). And even if logical truths are knowable *a priori*, this does not mean that they are not knowable *a posteriori* as well.

The reflective equilibrium method, moreover, fits better with the minimalistic, ontologically light account of normative truth that Parfit and Scanlon favour, according to which there are normative truths, but not robust or worldly normative

---

129. Specifically, Parfit (2011b) claims that “we cannot give helpful arguments for *Non-Contradiction* or *Modus Ponens*, nor could we have any direct reasons to believe these truths. These and some other logical truths are too fundamental to be supportable in such ways.” But nonetheless, “our belief in such truths can be fully justified” (508-509).

facts and properties. If normative truths do not have truthmakers, it makes more sense for the coherence of our normative beliefs (with each other and with our non-normative beliefs) to be sufficient for them to amount to knowledge, but it makes far less sense if our normative beliefs have to correspond to robust or worldly normative facts in order to be true. After all, under Robust Realism, our normative beliefs could in principle be fully coherent but completely out of touch with the (robust or worldly) normative facts. We have already seen that the robust form of metanormative realism is more plausible than the minimalistic form (Chapter 4). Since our responses are sometimes counterfactually dependent on the normative facts and properties, normative knowledge must involve some kind of interaction between us and these normative facts and properties. An adequate account of normative knowledge, therefore, must explain how this interaction is supposed to work, and a mere appeal to reflective equilibrium does not provide this explanation. As Scanlon suggests, normative judgements need to have “substantive merits” independent of reflective equilibrium itself for reflective equilibrium to even be relevant in the first place. If non-naturalism is true and these substantive merits, prior to reflective equilibrium itself, are not empirical in nature, then they clearly involve some kind of intuitive warrant that normative judgements must possess in order to be candidates for knowledge. So the non-naturalist needs an intuitionist account of normative knowledge, going beyond reflective equilibrium to confirm the *a priori* truth of certain fundamental normative judgements, while at the same time ruling out the possibility that these truths can also be known *a posteriori*.

#### 7.4. *Are (fundamental) normative truths conceptual truths?*

Take a fundamental normative proposition like the moral proposition that “[i]t is pro tanto wrong to engage in the recreational slaughter of a fellow person” (Cuneo and Shafer-Landau 2014, 405) or perhaps the instrumental principle that “there is a reason for *a*’s G-ing, or ... *a* has a reason to G, if G-ing would lead to the fulfilment of some desire or purpose or ideal that *a* now has, and *a* knows this” (Mackie 1990, 77),<sup>130</sup> or indeed the Agony Principle itself. If these propositions are conceptual truths—true by virtue of the concepts involved and the necessary relations between them—this would provide the non-naturalist with the intuitionist model of normative knowledge he needs. The intuitive warrant a normative judgement must possess in order to be a candidate for *a priori* knowledge could then be explained in terms of grasping the concepts involved in the judgement; to “intuit” that a normative judgement is true would be to grasp, or seem to grasp, its constituent concepts and the necessary relations between them. According to Parfit, some of our beliefs, including some of our normative beliefs, are “*self-evidently* true, in the sense that, if we fully understand these beliefs, we can recognize that they are, or must be, true.” If a belief is true entirely by virtue of the concepts involved in it and the relations between them, it is no wonder that we can know that this belief is (and indeed must be) true simply by fully understanding the content of this belief.

---

130. I am assuming for the sake of argument that this is a normative claim, rather than a non-normative, “hypothetical” reasons claim. See Chapter 2 (Section 2.3) for the distinction between normative and hypothetical reasons.

Parfit actually suggests that substantive normative truths “are self-evidently true, not because their truth is implied by the concepts they involve, but because it is so obvious that these claims are, or must be, true.” But he seems to think that conceptual truths are the same thing as analytic truths, and thus that substantive normative truths cannot be conceptual truths because they are not analytic (2011b, 508).<sup>131</sup> Conceptual truths, however, are not necessarily analytic: that nothing can be both red and green all over at the same time is a not an analytic truth, but it is nonetheless a conceptual one. And if fundamental normative truths are uniformly synthetic and (purely) *a priori*, and are known to be true solely because they seem obviously true when they are adequately understood,<sup>132</sup> then it would make sense if they had the same status as the truth that nothing can be both red and green all over (at the same time) and other conceptual truths.

Cuneo and Shafer-Landau outline four “marks” of a conceptual truth. The first is that “p is, if true, necessarily true,” the second, that “p enjoys framework status, fixing the boundaries as to what counts as a type of subject matter,” and the third, that “p’s denial would tend to promote bewilderment among those competent

---

131. The status of analytic normative truths is not something that is at issue between the naturalist and the non-naturalist. The naturalist can accept that analytic normative truths (such as the truth that murder is wrong, “murder” being defined as a wrongful killing) are purely *a priori*, just like non-normative analytic truths.

132. I say *seems* obviously true, because, as Parfit acknowledges, “when some belief seems to us self-evidently true, we may be mistaken. Such beliefs may be false.” Parfit, 508. The non-naturalist does not need our foundational normative beliefs to be infallible, only that, when some fundamental normative proposition seems to us to be obviously true if we fully understand this belief, this gives us reason to accept this proposition.

with its constituent concepts—a response to the effect that its denial would be almost crazy” (2014, 407-408).<sup>133</sup> If *p* exemplifies all of these marks, this may not *entail* that *p* is a conceptual truth, but it does give us good reason to believe that it is (408). The proposition that it is *pro tanto* wrong to engage in the recreational slaughter of a fellow person does seem to exemplify all these marks. If it is true, it is probably a necessary truth: it is not as if recreational slaughter just happens to be wrong in this particular world, and yet is somehow not wrong in other possible worlds. It probably “enjoys framework status” in the sense that it must be true unless there is something fundamentally wrong with the relevant discourse: if you engage in moral discourse but claim that it is not even *pro tanto* wrong to engage in the recreational slaughter of another person, you are probably not actually employing the moral concept “wrong” and are instead suffering from a conceptual confusion.<sup>134</sup>

Moreover, if you were to deny this proposition to anyone who engages in moral

---

133. The fourth mark is that “*p* is knowable a priori, simply by adequately understanding its constituent concepts and their relations to one another” (408). But whether or not fundamental normative truths are *a priori* is of course the very thing at issue here, and in any case, it is not clear why Cuneo and Shafer-Landau think that being knowable by adequately understanding its constituent concepts and their relations to one another is a *mark* of a conceptual truth, rather than just being what a conceptual truth is.

134. Cuneo and Shafer-Landau compare this to the situation with a proposition like “God is a perfect being.” This proposition is bound to be true unless there is something fundamentally wrong with the relevant discourse (theological discourse in this case, for instance because God does not exist), and if you engage in theological discourse (and therefore assume that there is nothing fundamentally wrong with it) but yet reject the proposition that God is a perfect being, this probably indicates that you are not actually using the theological concept “God” (408).

discourse and is competent with its concepts, your denial would indeed seem crazy. The same kind of things could be said about the instrumental principle: it is probably a necessary truth that, if G-ing would fulfil some goal of yours (and you know this), then you have at least some reason to G, and if you were to deny this to anyone who engages in the discourse of practical rationality, your denial would seem crazy and would probably indicate that you are not actually using the concept of a practical reason. And Parfit suggests something very similar about the Agony Principle: “[i]f we can have some reasons, nothing is clearer than the truth that, in the reason-implying sense, it is bad to be in agony” (2011a, 82). So there clearly is some reason to believe that fundamental normative propositions such as these are conceptual truths.

If fundamental normative truths are conceptual truths, this would also answer Copp’s main argument against normative non-naturalism: that no synthetic normative proposition can be strongly *a priori* (i.e. empirically indefeasible), because disagreement can always supply an epistemic defeater for any (synthetic) normative proposition. According to Copp (2007), “our warrant for believing a proposition can be undermined or weakened by the disagreement of others in cases in which we have no independent reason to think we are in a better epistemic position than they are” (48-49), and since disagreement is an empirical phenomenon, disagreement about a proposition always constitutes (at least some) empirical evidence against it. *Unless*, that is, disagreement “would not affect the credibility of the proposition to an ideal thinker—a thinker who had no psychological weaknesses or computational

limitations and who had a full conceptual repertoire.” And Copp suggests that disagreements about (synthetic) normative propositions are never due to psychological weaknesses or computational limitations or the lack of a full conceptual repertoire, such that these disagreements would not affect the credibility of these propositions to an ideal thinker.<sup>135</sup> His sole reason for holding this view, however, seems to be nothing but the fact that these propositions are synthetic, combined with the idea that, if the denial of a proposition could only issue from conceptual confusion, then this proposition must be analytic (51). Like Parfit, Copp makes the mistake of conflating conceptual truths with analytic truths.<sup>136</sup> Conceptual normative truths could be both synthetic and such as could not be rejected in the absence of a conceptual confusion. Disagreement about these propositions would not affect their credibility to an ideal thinker, who suffers from no conceptual confusions, and thus would not qualify as empirical evidence against them.

---

135. Copp actually only refers to *moral* propositions and *moral* disagreement, but his argument is just as relevant for normative propositions as a whole.

136. Cuneo and Shafer-Landau hold that

some conceptual truths have substantive content and needn't be obvious. The moral fixed points are hardly empty tautologies, but that doesn't distinguish them from many non-moral propositions plausibly viewed as conceptual truths. That justified true belief is insufficient for knowledge; that God's possible existence entails God's necessary existence; that meaningful statements may be neither empirically verifiable nor analytic—these are substantive truths, surely, and yet also good candidates for being conceptual ones. (2014, 408)

Conceptual, synthetic normative truths would remain “strongly *a priori*” (2007, 43) even in the face of substantial normative disagreement.<sup>137</sup>

I think the non-naturalist should say that fundamental normative truths—and all synthetic *a priori* truths for that matter—are conceptual truths. If a truth is both synthetic and *a priori*, it must tell us that at least two entirely separate properties are nonetheless necessarily connected in some way. And if the connection between these two separate properties is not something that we can access empirically, then our epistemic access to this connection must come from our access to our mental representations of these properties—that is, to our concepts. If our concepts adequately represent the relevant properties, they must represent the necessary connection between these properties, and thus there must be also a necessary connection between the concepts themselves. To fully understand a proposition, after all, is to fully understand its concepts and the relations between them, and thus, if I know that a proposition is true solely by understanding this proposition, then the concepts involved in this proposition must be sufficient for me to know that the proposition is true. I know that nothing can be both red and green all over (at the same time) because my concepts of red and green (and of space and time) are adequate mental representations of these properties, and the necessary relations between these mental representations—these concepts—thus reflect the necessary

---

137. A strongly *a priori* truth being one that is a) *weakly a priori*, which means that it can be reasonably believed in the absence of empirical evidence, and b) *empirically infeasible*, which means that it admits of no empirical evidence against it (42).

relations between the actual properties. Suppose I have adequate concepts of being (*pro tanto*) wrong and of slaughtering a fellow person for recreational purposes; if I know *a priori* that it is *pro tanto* wrong to engage in the recreational slaughter of a fellow person, then the adequacy of my concepts and my understanding of them must tell me that, whenever the concept of recreationally slaughtering a fellow person adequately represents something, the concept of being (*pro tanto*) wrong will also adequately represent this same thing, even though neither of these concepts is actually contained within the other.

Assuming that fundamental normative truths are conceptual truths, therefore, the question for the non-naturalist then becomes, “how do we know that our mental representations of the relevant properties are adequate?” My concept of being (*pro tanto*) wrong may be necessarily connected with my concept of recreationally slaughtering a fellow person, such that anything I represent with the first concept, I will also represent with the second (and must always do so unless my concepts change). But if this is to provide a basis for *a priori* knowledge, my concepts must be adequate representations of the relevant properties: the necessary connection between my concepts as mental representations must reflect a necessary connection between the properties they are supposed to represent. To know that my concepts of being wrong and of recreationally slaughtering a fellow person are adequate, given that these concepts are connected representationally, I must know that the properties themselves are necessarily connected in the way that my concepts represent them as being. And to know *this*, or at least to know that I know this, I must know that it

actually is *pro tanto* wrong to engage in the recreational slaughter of a fellow person. This does not mean that fundamental normative knowledge is not conceptual knowledge: even if I don't know that my concepts are adequate, they may nonetheless be so, and thus I may still have *a priori* knowledge of the relevant truth (assuming that you don't need to know that you know that *p* in order to know that *p*). But it does mean that just observing that fundamental normative truths must be conceptual truths if they are not empirical truths does not give us the complete epistemological account we are looking for.

#### 7.5. *Spontaneity and convergence*

For Scanlon, what makes a judgement a "considered judgement" (i.e. one that has the kind of "substantive merits" necessary for reflective equilibrium to provide justificatory force) is whether or not it is "something that seems to me to be clearly true *when I am thinking about the matter under good conditions for arriving at judgements of the kind in question*" (2014, 82). According to Skorupski (2010), the epistemic basis of normative knowledge—and thus what constitutes good conditions for arriving at normative judgements—"consists solely in spontaneity and convergence" (405). "*Spontaneity*," he explains, "is a property of responses and dispositions to respond. The basic idea is that a spontaneous response or disposition is one that comes in the right way from, is genuinely that of, the actor." This is as opposed to "one that is accepted uncritically into one's thinking from others, or one that results merely from a wish to please or to annoy ... and so on"; one that is the result of what Kant would

call “subjectively determining grounds” or “alien causes.” A truly spontaneous response, on the other hand, “is simply one that is free of such grounds or causes” and which instead “comes in the right way from the actor: that is, from the actor’s nature” (406-407). What determines whether or not a normative judgement is warranted is precisely whether or not it is spontaneous. As Skorupski puts it, in the form of what he calls the “*norm of spontaneity*”:

when one is warranted in taking oneself to have a spontaneous disposition to judge that a set of facts  $\pi_i$  would give one reason to  $\psi$  then, in the absence of defeaters, that warrants the judgement that the set of facts  $\pi_i$  would give one reason to  $\psi$ . (409)

Take a normative judgement like the Agony Principle. By Skorupski’s norm of spontaneity, to find out whether or not we are warranted in taking this judgment to be a genuine *a priori* truth, we must ask, first of all, whether or not we really are spontaneously disposed to make this judgement. For Skorupski, “spontaneity in one’s cognitive, affective or practical dispositions is typically marked by a certain experienced or felt normative harmony.” Normative harmony occurs when a “disposition to  $\psi$  ... blends with a disposition to take oneself (more or less explicitly) to have reason to  $\psi$ .” And a disposition to take oneself to have reason to  $\psi$  “is an *experience, impression or spontaneously persuasive representation of a given response—belief, feeling, or action—as reason-supported: normatively apt, proper*” (407). Therefore, if we have the experience or impression that wanting to avoid agony is reason-supported or normatively proper, and this disposition blends with a further

disposition to actually want to avoid agony,<sup>138</sup> and in a way that produces an experience of harmony between the two dispositions, then we have warrant that the Agony Principle is indeed a spontaneous judgement on our part, and thus that it expresses a genuine *a priori* truth.

By the norm of spontaneity, however, this is not all that is required for a normative judgement to be warranted; there must also be an absence of “defeaters”: external or “alien” causes of the judgement that would explain away its apparent spontaneity and warrant. As has been mentioned, if a response is produced by a desire to please or annoy, for instance, then it is not a truly spontaneous response and therefore is not a warranted response for the actor to make, or so Skorupski argues. But why does the fact that a response can be explained by such factors mean that it is not warranted, whereas truly spontaneous responses universally (and for that very reason) do possess warrant? On this account,

explanations that subvert or explain away the spontaneity of a response (or lack of a spontaneous response) also remove its epistemic value. One could say that they work by showing that the response isn't, so to speak, tuned solely to its object (the object of normative assessment). It is not solely an interaction between the object and the subject's nature that gives rise to the response. There is interference from other factors. (409)

What makes a spontaneous response also a warranted response is precisely the fact that such responses must, by virtue of the absence of any interfering causes (such as

---

138. By contrast, in cases where it seems to me “that I have reason to  $\psi$  and yet I find I'm not at all disposed to  $\psi$  ... whether I really do have reason to  $\psi$  comes into question” (408).

ulterior motives on the part of the actor) be based solely on the normatively relevant features of the object.<sup>139</sup>

Imagine a response that we are generally disposed to make, but for which we could find no subverting explanation: one that could not plausibly be explained away by appealing to factors that clearly have nothing in particular to do with whatever normatively relevant features the object of the response might possess. Now suppose that we ask the question, "Why are we so disposed to make this response, to respond in this particular way to that particular object?" By the hypothesis, the answer to this question could not be an epistemically negative one that shows the response to be lacking in warrant. Could it be an epistemically neutral one: one that indicates neither that the response is warranted nor that it is unwarranted? I don't think it could. A subverting explanation for the response would of course be one that shows it to be a product of factors that are not among the normatively relevant features of the object. If such an explanation from normatively *irrelevant* factors is not available, the only alternative answer to the question, "Why do we respond in this way?" must find its answer in normatively *relevant* factors. These two alternatives, it would appear, are exhaustive: if a response or disposition to respond in a certain way to an object is not the product of normatively irrelevant factors (rather than the normatively relevant features of the object), then the reason (or reasons) for this response must be a normative reason (or

---

139. The object here is simply whatever it is that the response is a response to, and hence it might be a set of facts rather than a specific concrete object.

reasons), and hence it must be precisely the normatively relevant features of the object that the response is a response to.<sup>140</sup>

The story is still not complete. The final ingredient needed for a normative judgement to be warranted is for it to test positively against the normative responses of others. Just like the kinds of normatively irrelevant causes that can defeat the warrant one might appear to have for a normative judgement, disagreement can supply a defeater for one's judgement in cases "when one has insufficient reason to believe that the disagreeing judgements are faulty." As such, Skorupski advocates what he calls the "*Convergence thesis*":

if I judge that *p*, I am rationally committed to holding that *either* inquirers who scrutinized any relevant evidence and argument available to them would agree that *p* or I could fault their pure judgements about reasons or their evidence. (412)

This follows from the fact that, if I make a truly spontaneous normative judgement, and my judgement is therefore not the product of normatively irrelevant causes but instead is a direct response to the normatively relevant features of the object, then

---

140. Of course there is a third option: that the response is neither the product of normatively irrelevant factors, nor a warranted response to normative reasons—it could be a completely random and arbitrary response. But if a response is neither the product of purely naturalistic causes, nor a response to (non-natural) normative reasons, we would not expect the kind of uniformity we see in our responses to something like agony; we would instead expect a variety of responses to it, some people wanting to avoid it, others wanting to seek it out. The uniformity we do in fact see is surely something that prompts a "why?" question, whether this is to be answered by some kind of purely causal, non-normative explanation or by a normative reason. And in any case, we are taking ourselves to have established metanormative realism here (by the arguments of Chapters 3 and 4). Since normative reasons, facts and properties, I believe, do exist, they must find their place somewhere, and it would not be surprising if they found their place precisely where non-normative explanations fail to account for our responses.

anyone else who likewise responds to the object without being influenced by normatively irrelevant causes will thereby also be responding directly to these same normatively relevant features, and will thus come to the same normative judgement. If they do not come to the same judgement, then at least one of us must be unduly influenced by normatively irrelevant factors; if I do not have sufficient reason to believe that the other party is compromised in this way, I also do not have sufficient reason to believe that I am not compromised in a similar way, and thus I am not going to be warranted in my original judgement. If a normative judgement is warranted if and only if it is a response solely to the normatively relevant features of its object, then, since the normatively relevant features of the object will be the same no matter who is making the judgement, the only way that our normative judgments could fail to converge is if at least one of us is subject to normatively irrelevant influences. The absence of convergence on a normative judgement is therefore evidence that the judgement is not truly spontaneous but a product of compromising influences. If there is widespread convergence on a judgement, on the other hand, this constitutes (at least) an absence of evidence that the judgement is compromised, which, when combined with the other factors mentioned above—the felt harmony between an impression that a certain response is or would be reason-supported or normatively apt, and a disposition to respond in that very way, plus the absence of any plausible non-normative explanation for the judgement that would remove its warrant—will suggest that the judgement is indeed fully warranted.<sup>141</sup>

---

141. Even if there is full convergence on a normative judgement, this

How would this model of normative knowledge work when applied, once again, to Parfit's Agony Principle? Remember, for Skorupski, a response is truly spontaneous only if it is "solely an interaction between the object and the subject's nature that gives rise to the response" (409), and as such there is no interference from other (normatively irrelevant) factors. For the Agony Principle to be spontaneous, interaction between our nature and the object of the judgement (agony) must give rise to 1) a disposition to make the judgement, that is, an experience or impression to the effect that wanting to avoid agony is a reason-supported or normatively apt response to the object, and 2) a disposition to (always) want to avoid agony, which is attendant upon, and experienced as blending harmoniously with, the first (normative) disposition. If these conditions are met, the Agony Principle is backed up by a first-person experience of normative harmony, given rise to, at least in part, by interaction between the subject's nature and the nature of the relevant object, making it a candidate for a warranted normative judgement. And indeed, these conditions very much seem to be met as far as Parfit is concerned: as he says,

---

judgement might still fail to be warranted, because this convergence could itself be the product of distorting influences (e.g. evolutionary factors). But Skorupski is not saying that convergence is a sufficient condition for a normative judgement to be warranted, only that it is a necessary condition. The absence of convergence undermines the warrant for a normative judgement, but the presence of convergence, while necessary for the judgement to be warranted, means nothing in the absence of spontaneity; and if the convergence for a judgement is the product of distorting influences, the judgement is not going to be spontaneous. A convergent normative judgement that is nothing but the product of evolutionary factors (which apply to all of us and which are therefore the reason for our convergence on this judgement) would not be a spontaneous judgement in Skorupski's sense of "spontaneous."

remembering what it is like to have intensely painful sensations makes it hard for him to believe that, “though the approaching flames threaten to cause me excruciating pain, this fact does not count in favour of my wanting ... to move my hand away,” even if he happened not to have any goals or desires that would be furthered by moving his hand away (2011a, 81). I think we can also assume that Parfit (like most of us) always does have a desire to avoid agony, a desire which blends harmoniously with his distinct impression that this particular response to agony is reason-supported.

So, for Parfit at least, the Agony Principle seems to have the kind of first-personal support necessary, on this account, for it to be a warranted normative judgement. The next step is to ask whether or not there is anything that might explain away this first-personal support—this appearance of spontaneity and normative harmony. Might his judgement simply be the product of normatively irrelevant causes, rather than flowing spontaneously from a pure interaction between Parfit and the object of his judgement (agony)? But what, other than the nature of agony itself, might supply the reason why Parfit thinks he has a (non-instrumental) reason to want to avoid agony? One might think that Parfit just really doesn't like agony, and wants others not to do anything that might cause him agony, and this somehow provides an ulterior motive for him to think and perpetuate the idea that his desire to avoid agony is reason-supported. Perhaps if others think that this desire is reason-supported, and not just some (normatively) arbitrary preference of his, they will be more inclined to respect this desire, and hence less likely to do

anything that might cause him agony, and perhaps more likely to help him avoid agony in the future. But this would not really explain why he is disposed to respond to agony by wanting to avoid it in the first place, nor why this response actually feels reason-supported (to him). And in any case, it is not clear why perpetuating the idea that one's desire to avoid agony is reason-supported would help one to avoid agony; to say that one always has a reason to want to avoid agony is not to say that others have sufficient reason to spare you agony. Not only that, but why would Parfit even dislike agony in the first place, if not through an interaction between his nature as a subject who knows what agony is like and the nature of agony itself? In any case, it is not obvious that there is a debunking explanation available that would prevent Parfit's judgement from being warranted.

Finally, we must ask how the Agony Principle converges with the normative judgements of others. As Parfit acknowledges, some people appear to reject it, believing instead that "the nature of agony gives us no reason to want to avoid being in agony" (81). These people are those who endorse what he calls "Subjectivism about Reasons" (58): the position that the only reasons we have are to respond in ways that fulfil (at least some of) our desires (from which it follows that we have reason to avoid or want to avoid agony, not because of the nature of agony, but only because being in agony would conflict with our desires).<sup>142</sup> And since, according to Parfit, "Subjectivism about Reasons is now very widely accepted" (65), this may

---

142. This is the same thing as what we referred to as *Internalism* about reasons in Chapter 3 (Section 3.5).

seem to indicate that the Agony Principle simply does not pass the convergence test, and thus that it does not express a warranted, *a priori* truth after all. But some of these subjectivists no doubt reject the Agony Principle, not because they find another conflicting first-order normative judgement more plausible, but for metaethical or metanormative reasons, perhaps because they think that reasons which do not fundamentally stem from our desires, or reasons to have certain desires as such (like a reason to want to avoid, and to try to avoid, agony even if you don't in fact want to avoid agony), are metaphysically or epistemologically queer. Assuming, for instance, that the Argument from Queerness fails against these kinds of reasons,<sup>143</sup> the disagreement of these subjectivists will not undermine the warrant for Parfit's judgement. Other subjectivists, however, do seem to be making a conflicting normative judgement of their own, which might be just as spontaneous for them as Parfit's judgement appears to be for him; "Korsgaard for example writes, if some act 'is a means to getting what you want ... no one doubts that this is a reason'. Williams writes: 'Desiring to do something is of course a reason for doing it'" (65). It follows from these judgements that if (for some reason) one desired to be in agony, one would have a reason to pursue agony rather than to avoid it, and these normative judgements appear to conflict, rather than converge, with the Agony Principle.

However, this appearance, I think, is an illusion: the Agony Principle only says that we all have *a* reason to want to avoid agony, not that we always have

---

143. We of course rejected the Argument from Queerness against non-natural normative reasons, facts and properties in Chapter 5.

*decisive* or *sufficient* reason to do so. Having a reason to want to avoid, and to try to avoid, agony is compatible with also having a reason, perhaps even sufficient reason, to do things that involve being in agony (resisting torture, for instance), and perhaps with actually wanting to be in agony (maybe so one can endure it for Nietzschean purposes). And if wanting to do something always provides a reason for doing it, even if it would involve being in agony, as Korsgaard and Williams both firmly believe, this does not mean that we do not also have a reason to (want to) avoid agony; we could have conflicting reasons both to do and not do something that would fulfil some of our desires but would also involve being in agony. So the Agony Principle is not in fact inconsistent with the judgements of subjectivists about reasons like Korsgaard and Williams, and hence these people only reject the Agony Principle on purely metaethical/metanormative grounds, something which does not indicate that this first-order normative judgement fails to converge with the first-order normative judgements of others who are equally rational.

Would anyone be spontaneously disposed to reject the Agony Principle, so long as we make explicit the qualification that we may also have other reasons not to avoid agony (or at any rate to do things which may require being in agony)? Safe to say, we are all disposed to want to avoid agony; the masochist who desires pain, I believe, at most desires pain up to a certain intensity and not beyond—such a person is likewise disposed to want to avoid a certain level of agony. And even someone who genuinely desires to be in agony probably also has a desire to avoid it, even if this desire is outweighed by the strength of his desires that, for whatever reason,

bend towards being in agony.<sup>144</sup> We know that, for Parfit, this response to agony feels reason-supported or normatively appropriate, hence why he is disposed to endorse the Agony Principle. Does our desire to avoid agony universally feel reason-supported, that is, like a response that is well attuned to its object, without interference from any distorting influences? I think the answer is yes: for those who know what agony is like, the desire to avoid it flows directly from this knowledge—knowledge based on direct experience of the object and its nature. We experience agony, we know what it's like, and so we want to avoid it, and since this desire cannot be separated from our knowledge of the nature of agony (and indeed, it seems that it would vanish entirely in the absence of this knowledge), I think we can say that wanting to avoid agony is a response that is well attuned to its object if any response is, and thus, to beings like us who experience agony and know what it's like, having some desire to avoid agony genuinely feels like a reason-supported or normatively appropriate response. When the Agony Principle is properly understood, therefore, and when all (for the present purposes) irrelevant metaethical/metanormative doubts are set aside, I think we find that it does converge well with the normative responses of others, and thus that it is a truly spontaneous response to its object (if there are any such responses). As this

---

144. An example of this might be someone who is experiencing extreme levels of guilt and/or self-hatred, and who perhaps believes that they deserve to be severely punished and suffer for some transgression (real or imagined). Such a person might actually want to experience a degree of agony, but it seems that the reason for this would be, at least in part, to mitigate the psychological/emotional agony the person is currently experiencing, and thus this person, despite having a desire to be in agony, also clearly has a desire to avoid it.

illustrates, the notions of spontaneity and convergence do seem to provide a way for us to know certain fundamental normative truths *a priori*, and to differentiate genuine *a priori* truths from merely (epistemically) possible or apparent ones.<sup>145</sup> If our (fundamental) normative judgements conflict, we can potentially use the criteria of spontaneity and convergence to settle our disagreements.

### 7.6. *The problem with spontaneity*

As Skorupski understands it, spontaneity contrasts fundamentally with “receptivity” as a basis for knowledge. Knowledge of the natural world “involves productivity in the object, on the one hand, and receptivity in the knowing subject, on the other,” whereas normative knowledge, on the other hand, “involves no receptivity at all” (2010, 405). What this means is that normative knowledge, unlike empirical knowledge, is not causal in nature; when we know something about the natural world, this is because we have been causally affected (through our *receptive* faculties, i.e. our senses) by some part of it,<sup>146</sup> but when we make a warranted

---

145. This account could be taken either as a way for reflective equilibrium to get off the ground, or as a foundationalist model of normative knowledge in its own right. Spontaneity and convergence are clearly supposed to be foundations for normative knowledge (they constitute its “epistemic basis” as Skorupski puts it (2010, 405)). But they could also supply the “substantive merits” of the judgements we make in reflective equilibrium (Scanlon 2014, 82), in cases when that process leads us to arrive at epistemically justified normative conclusions.

146. Skorupski specifically says that “[k]nowledge of reasons relations [i.e. normative knowledge] does not require either identity with or causal relation to substantial facts” (405). And by a “substantial fact” he means a fact that can cause and/or be caused (404).

normative judgement, not only is this judgement not the product of “alien causes” or normatively irrelevant causes—it is not the product of any causal influences whatsoever. Instead, as we have mentioned, if a response is truly spontaneous and thus normatively warranted, it must be “solely an interaction between the object and the subject's nature that gives rise to the response” (409). On this account, therefore, normative knowledge must arise solely from these kinds of non-causal interactions (between the nature of the object of knowledge and the nature of the knowing subject).

Assuming that the Agony Principle is a fundamental normative truth, if this truth is known *a priori*, this must be because, when we respond to agony by having a desire to avoid it, it is nothing but the nature of agony itself that gives rise to this response in us, and in an entirely non-causal manner at that. But if the interaction between our nature and the object of our response (when we respond to agony by wanting to avoid it) is not causal interaction, what kind of interaction is it? Is the non-naturalist really committed to saying that agony does not cause us to want to avoid it, that when we experience agony and consequently want to avoid it, this is not a causal response to agony? Undoubtedly, my *experience* of agony (or the fact that I've experienced agony) does cause me to want to avoid being in agony: if I had never experienced agony, I would not have any desire to avoid it.<sup>147</sup> But this might not mean that agony itself, or the nature of agony, exerts any causal influence on me.

---

147. Or at least if I had never experienced anything else, such as less intense pain, that might indicate to me what agony is like.

If my desire to avoid agony is a truly spontaneous response, then it is a response solely to the nature of agony. The non-naturalist would argue that we cannot say that, if agony had a different nature than it actually has, I wouldn't respond to it in this same way: if agony had a different nature (if *what it's like* to be in agony was different), it would not really be agony at all. There is no possible world in which the nature of agony (*what it's like* to be in agony) is different, so there is no possible world in which I respond differently to agony (assuming that I am responding solely to the nature of agony). There is no counterfactual dependence between my response to agony and its nature, so by the counterfactual theory of causation, the nature of agony does not cause me to respond to it in the way that I nonetheless do.

The basic idea, therefore, is that some of our responses, and some of our knowledge, is counterfactually dependent, and thus fundamentally causal in nature—these are our receptive responses. Other responses of ours, however, are responses to things (namely, the inherent natures or essences of things, such as the nature of agony) that necessarily are the way that they are and which, for this reason, do not support counterfactuals, and these are our spontaneous responses. Our spontaneous responses do not admit of any causal explanation; the reason for a spontaneous response—why we respond in that particular way—cannot be a causal or explanatory reason, so it must be a normative reason, and this is why our spontaneous responses are warranted and thus constitute the basis for normative knowledge. This also appears to tell us that (fundamental) normative knowledge is always *a priori* and never *a posteriori*: since there are two fundamental kinds of

knowledge—the receptive and the spontaneous—and empirical knowledge plainly falls into the receptive category (involving as it does the causal stimulation of our sensory receptors), whereas normative knowledge falls into the spontaneous category, normative knowledge cannot possibly be empirical knowledge. And this obviously means that normative naturalism must be false. This is made especially plausible when we consider that, if we succeed in finding a fully causal, naturalistic explanation for a response (a desire, say), this seems to remove any need to posit a normative reason for that response, and Occam's Razor would seem to count against doing so.

The problem with this line of reasoning, however, and hence the non-naturalist's position as a whole, is that, if the reason why we respond in the way that we do to something like agony is the inherent nature of agony itself, then the nature of agony must be able to explain why we respond to agony in that specific way. That we universally want to avoid and try to avoid agony is after all an empirical observation, and thus we can reasonably ask what explains this observation. And if the answer comes down to nothing but the nature of agony itself, and this is supposed to indicate that our response is not the product of normatively irrelevant causes but is instead truly spontaneous and warranted, then it must be the case that agony, by its nature, possesses some normative property, and that it is this property that ultimately explains why we respond to agony in the way that we do.<sup>148</sup> But this

---

148. Skorupski's view implies that there are two ways in which we respond to agony: 1) by wanting (or having a disposition to want) to avoid it, and 2) by taking

surely means that we can know that agony has this normative property on the basis of IBE, and thus that this piece of fundamental normative knowledge can be acquired empirically and not just *a priori*. Someone who had never experienced agony themselves (or anything that might suggest to them what agony is like) could nonetheless observe that those who *have* experienced agony universally want to avoid it, ask what explains this empirical observation, find no plausible non-normative, purely causal explanation and hence conclude that the best explanation for this observation is that agony has some normative property. And since the non-naturalist is saying that there is no adequate causal explanation for the fact that those who have experienced agony universally want to avoid it, he must admit that this would be a sound conclusion to draw for someone who has never experienced agony but who is trying to explain why those who have universally want to avoid it. But this would mean that fundamental normative knowledge can be acquired using observation and IBE, without the need for a spontaneous response to the object on

---

ourselves (or having a disposition to take ourselves) to have a reason to want to avoid it. For normative harmony to be achieved, and thus for the Agony Principle to be warranted, both of these dispositions must be present. But the non-naturalist could argue that it is only the second response/disposition that is normative and spontaneous, whereas the first response/disposition is indeed purely causal (these two ways in which we respond to agony corresponding to the naturalistic, causal order and the non-natural, normative order respectively), and thus avoid having to say that we don't respond to agony in a causal manner, or that our desire to avoid agony is not explained by some property of agony (normative or otherwise). Yet since the Agony Principle specifically claims that our desire to avoid agony is reason-supported (not just our disposition to take ourselves to have this reason), and since, for the non-naturalist, a response must be spontaneous if it is reason-supported or warranted, it is not enough for only the second response to be spontaneous: the first response—the actual desire itself—must be spontaneous as well, or else the Agony Principle will not only be unwarranted but false.

the part of the knowing subject—and this is all the naturalist requires and the very thing the non-naturalist thinks is impossible.

The non-naturalist might still deny that the nature of agony explains why we respond to agony in the way that we do, insisting instead that explanation is simply not the right relation between the nature of agony and our response to it. The nature of agony, for the non-naturalist, does not *explain* why we respond to agony by wanting to avoid it; it *gives us a reason* for this response, and the domain of explanations is fundamentally distinct from the domain of reasons. The non-naturalist might also accuse the naturalist of an equivocation: there is a sense, even for the non-naturalist, in which the nature of agony can be said to “explain” why we respond to agony as we do, because it answers the question, “Why do we respond to agony in the way that we do?” But this does not mean that the nature of agony *causally* explains our response to agony, and to assume that it does based upon the above question alone is to equivocate between two sense of “explain”: the narrower, causal sense and the broader sense that involves answering any kind of “why?” question whatsoever, whether or not this question is specifically causal.<sup>149</sup> Yet this response on the part of the non-naturalist does not seem to eliminate the possibility that someone who has never experienced agony and has no idea what it’s like could nonetheless arrive at the conclusion that there is a universal reason to want avoid it by inference to the best explanation from the empirical fact that those who do know

---

149. The question, “Why does  $\langle 2 + 2 = 4 \rangle$ ?” is presumably not a causal question, and so the answer will not constitute a causal explanation, but it is clearly still possible to explain (in the broad sense) why  $\langle 2 + 2 = 4 \rangle$ .

what its's like universally want to avoid it. If such a person could not, in principle, make this inference, this would have to be either because there is some debunking causal explanation for why we want to avoid agony which shows it not to be a warranted response on the non-naturalist's own terms, or because the absence of such a causal explanation does not in fact suggest that there must be a normative reason for this response after all. The whole point of spontaneity as a basis for normative knowledge, however, is that, if there is no causal explanation for a particular response, there must be a normative reason for that response, and that it is for this reason and no other that the response occurs.

Someone who does not possess the concept of a normative reason, or who does not know that there are any normative reasons, would not be able to conclude from the universal desire to avoid agony, plus the absence of a plausible debunking causal explanation for it, that there is a normative reason to want to avoid agony. But someone who does know that there are normative reasons must also know that, if we have sufficient reason to respond in a certain way, and we are subject to no distorting influences that might prevent us from responding as we have sufficient reason to respond, then we would indeed respond in that particular way. And if there is no debunking causal explanation for why we respond to agony by wanting to avoid it, and thus no distorting influences that might makes us want to avoid agony even though we have no reason to do so, then, if the nature of agony does give us a reason to want to avoid agony, we would expect there to be a universal desire to avoid agony—and this is indeed what we find. On other hand, given that

(as the non-naturalist assumes) there is nothing causing us to want to avoid agony, if there were no normative reason for us to want to avoid agony, we would not expect there to be a universal desire to avoid agony. The fact that there is this universal desire, therefore, is empirical evidence for the hypothesis that there is a normative reason to want to avoid agony. To arrive at this conclusion (by IBE), one does not need to have a spontaneous response to agony oneself: one just needs to know 1) that there are normative reasons, 2) conscious beings universally respond to agony by wanting to avoid it, and 3) that there is no (plausible) debunking causal explanation for this response. This appears to make this fundamental normative truth also an empirical, naturalistic truth, even if we grant the non-naturalist's premise that truly warranted responses are not causal.

It is difficult to see how, on the spontaneity and convergence model of normative knowledge, it can be denied that fundamental normative truths can, in principle, be acquired *a posteriori*, even if these truths can always also be known *a priori*. If it is possible to know a normative truth *a priori* on the basis of one's own spontaneous response and how well it converges with the responses of others, it should be possible to know this same truth empirically by observing the spontaneous responses of others (and how well these responses converge with each other). It might be asked how one could possibly know that another person's response is truly spontaneous if one does not have that spontaneous response oneself, and thus the felt experience of normative harmony that comes with it. But if one knows that another person's response is not causal—that there is (probably) no

causal explanation for it (which is what the non-naturalist claims anyway)—and that there is also widespread convergence on this response, and therefore that there is likely to be at least some kind of reason for it, then one can infer that this reason must be a normative reason, and one can thus arrive at knowledge of a fundamental normative truth by empirical observation and IBE alone without relying on a form of *a priori* intuition; we do not necessarily need to be spontaneously disposed to make a normative judgement ourselves in order to know a (fundamental) normative truth.

The non-naturalist might argue that this is just a specific case of a quite general phenomenon: that *a priori* truths can also be known *a posteriori* by observing reliable sources of information about those truths.<sup>150</sup> The truths of arithmetic, for instance, can be known *a priori* by working them out oneself, or *a posteriori* by using a calculator. Andrew Wiles knows that Fermat's Last Theorem is true *a priori*, but I know that it is true (I imagine) *a posteriori*. If I can acquire (fundamental) normative knowledge *a priori* based on my own spontaneous responses, therefore, I can also acquire this same knowledge *a posteriori* based on observing the responses of others. For me to be able to acquire normative knowledge from my own responses, I must have reason to believe that I am in good conditions for responding in the appropriate manner, and if am not in good conditions (if I don't know what agony is like, say), I might nonetheless have reason to believe that others are in such conditions, in which case their responses, and my observations of them, will

---

150. Yudkowsky (2007b) even claims (perhaps hyperbolically) that “[t]here is nothing you can know ‘a priori’, which you could not know with equal validity by observing the chemical release of neurotransmitters within some outside brain.”

constitute a sound basis for normative knowledge (so long as my own responses ever constitute a sound basis for normative knowledge). But this does not differentiate the normative from any other domain in which *a priori* knowledge is possible, such as logic or mathematics. And if these domains are non-natural, in spite of the fact that the fundamental truths within them can also be known *a posteriori*, then the normative domain, the non-naturalist might argue, is still non-natural as well. The problem with this response is that (as we saw in Chapters 4 and 5) logic and mathematics are simply not best construed as non-natural; although logical and mathematical truths probably can be known *a priori* without relying on IBE, these domains are nonetheless perfectly continuous with science.

In Chapter 4 (Section 4.6), I argued that normative non-naturalism amounts to a kind of epiphenomenalism about the normative, and that epiphenomenalism about the normative is implausible for the basically same reason that epiphenomenalism about conscious experience is implausible. Phenomenal properties appear to support counterfactuals: if our conscious experiences had different phenomenal properties, we would sometimes respond differently to them as a result. If the experience I have when I place my hand on a hot stove had the phenomenal property of being pleasant, rather than the phenomenal property of being painful, I might keep my hand on the stove for a period rather than instantly withdrawing it. And this, I believe, is what happens when we respond to agony by wanting to avoid it: we have a conscious experience, and we respond to that experience in a certain way by virtue of its phenomenal properties. Agony is simply an experience which has the

phenomenal property of being extremely unpleasant. If certain experiences did not have this property, but instead had the property of being pleasant, say, we would not respond to those experiences by wanting to avoid them. This means that the phenomenal property that makes agony what it is—the property of being extremely unpleasant—supports counterfactuals and thus has causal standing. For the non-naturalist, this means that our universal response to agony must not in fact be warranted after all, because it is a caused response and therefore not a truly spontaneous one. But just because our response to agony is caused by a phenomenal property surely does not mean that this response is not warranted. Given that it is this phenomenal property that causes us to want to avoid agony (or experiences to which we give the label “agony”), we can ask why this phenomenal property causes us to respond in this particular way. And if we can answer that the best explanation for this observation is that being extremely unpleasant gives us a normative reason to want to avoid experiences which have this property, then it is not spontaneity, as the non-naturalist understand it, that provides the epistemic basis for (fundamental) normative knowledge, but inference to the best explanation.

### *7.7. Is normativity really the best explanation for our response to agony?*

I have argued that our aversion to agony is not plausibly construed as a non-causal response. Although this does not entail that this response is not warranted, it might remove some of the rationale for thinking that it is warranted in the first place. If wanting to avoid agony is just a straightforward causal response to a certain

natural property (the property of being extremely unpleasant), the non-naturalist could argue that there would not be any reason to think that this response is warranted. The non-naturalist's argument, remember, is that, if a particular response is not a causal response—that there is probably no “debunking” causal explanation for it—and if there is enough convergence on that response to suggest that there is probably at least some reason of some kind for it (whether causal or normative), then since, by the hypothesis, there is (probably) not a causal reason for this response, the reason that there is for it must be a normative reason. So if, on the contrary, there *is* a causal reason for our response to agony, in the form of the natural fact that we want to avoid experiences that are extremely unpleasant, then what reason is there to think that this response is warranted, and not just another response with a “debunking” causal explanation for it?

For a naturalist such as myself, the best explanation for the fact that we want to avoid experiences which are extremely unpleasant is that experiences which have this phenomenal property also, and for that very reason, have the normative property of giving us a reason to want to avoid them.<sup>151</sup> But why think that this normative explanation is really the best explanation for the uncontroversially natural fact that we want to avoid agony (or that we want to avoid extremely unpleasant experiences)? Wouldn't another uncontroversially natural fact represent a superior

---

151. Or that they have the property, *ought-to-be-avoidedness*. I will presently leave it open whether the normative property is reducible or identical to the phenomenal property, or whether it is *sui generis* (metaphysically, not epistemologically). We will discuss reductionism in the next (and final) chapter of this dissertation.

explanation, perhaps because it would be more informative and/or more parsimonious? Or one could even argue that there actually isn't anything to explain here, on the grounds that it is simply a brute fact that we want to avoid experiences which we find extremely unpleasant. Perhaps it is simply the phenomenal property itself which explains our response to these experiences, and there is simply no need to explain why we respond as we do to that phenomenal property. After all, if it is a normative property that explains why we respond in that way to experiences with that phenomenal property, then the link between that normative property and our response would arguably be just as brute and inexplicable, so why not just accept the bruteness and inexplicability at the phenomenal level, rather than positing a normative property as a supposed explanation which really only succeeds in pushing the mystery back?

First of all, I think there is genuinely something to be explained here: a universal desire to avoid extremely pleasant experiences and to pursue extremely unpleasant experiences is at least *conceivable*.<sup>152</sup> Nature might conceivably have disposed us to avoid stimuli that give us pleasure and to pursue stimuli which cause us pain and suffering.<sup>153</sup> Yet this is of course the reverse of what we actually find,

---

152. Not to say that this combination is *possible*. But if it isn't possible, this is itself something that calls out for explanation.

153. The aversion to pain is probably not completely universal, and thus it is not the aversion to *pain* as such that is the basis for the conclusion that we have a normative reason to want to avoid agony. Kahane (2010) points out that

and thus we can reasonably ask why this is the case. Pleasant experiences have an ineffable phenomenal character that differentiates them essentially from unpleasant experiences, which have a different ineffable phenomenal character, and we find that the specific phenomenal characters of these different experiences universally give rise to certain specific responses on the part of those who have these experiences: attraction in response to the pleasant experiences, aversion in response to the unpleasant experiences. So the question is why do those particular conscious experiences, with those particular ineffable phenomenal characters, universally give rise to those particular responses in us: why is it that we want and try to avoid unpleasant experiences such as agony—experiences with *that* ineffable phenomenal character—and want and try to pursue pleasant experiences—experiences with *that*

---

there seem to be cases, such as when patients undergo frontal lobotomy, where the sensation of pain seems to lose its hurtfulness. When lobotomy patients can still feel the sensation of pain but they do not dislike or mind it, and there seems to be no reason to suppose that what they experience is nevertheless bad. (28)

These are cases in which pain is experienced, but without suffering. Kahane himself thinks that “the state that is intrinsically bad is not that of having a sensation of pain, but that of suffering—of having this sensation and disliking it. And what makes *this* state bad is what it feels like” (38). The aversion to *suffering*, I believe, is universal, and agony is clearly a form of intense suffering. Cases of *pain asymbolia*, in which subjects feel pain but appear to be indifferent to it, are not instances in which the aversion to suffering is not present, because these are cases in which pain is not experienced as unpleasant, so these cases do not break the universal link between unpleasant experiences and the aversion response. One might think that the fact that suffering involves disliking a certain sensation removes the need for an explanation of our aversion to it. Yet the experience of disliking a certain sensation is itself a phenomenal sensation, so there is still a need to explain why we respond in that particular way to that particular sensation. To not like a certain sensation is surely just to experience that sensation as unpleasant (and thus to have an unpleasant experience).

*other* ineffable phenomenal character—rather than wanting and trying to avoid pleasant experiences and wanting and trying to pursue unpleasant experiences instead?

Nor can the link between the phenomenal characters of these experiences and the different responses to which they give rise just as plausibly be taken to be as brute as the link between our responses and our normative reasons. As a conceptual matter, if we have sufficient reason to respond in a certain way, and we are subject to no distorting influences that might prevent us from responding as we have reason to respond, this surely means that we will respond in that way; the link between having a reason to respond in a certain way and actually responding in that way is not something that calls out for explanation, beyond appealing to the absence of any distorting influences that would compromise this link.<sup>154</sup> If we want to explain why someone did what they had sufficient reason to do, for instance, it seems to be enough to cite the absence of distorting influences (like depression, weakness of will, or excessive emotion, etc.), either explicitly or implicitly by appealing to the person's character (someone who does what they have sufficient reason to do by virtue of their character is surely just someone who does not tend to be affected by these distorting influences). The link between our phenomenal experiences and our responses to them, on the other hand, does seem to call out for explanation: it is not a

---

154. Unless, that is, we possess some form of radical libertarian free will, whereby we can have sufficient reason to respond in a certain way and be subject to no distorting influences, yet nonetheless just arbitrarily choose not to respond in that way. I will assume for the sake of argument that this is not the case.

conceptual truth that we want to avoid unpleasant experiences but not pleasant ones (so long as we are subject to no distorting influences). It makes sense, therefore, for the explanatory buck to stop at our responses to our normative reasons, rather than at our responses to our phenomenal experiences, and thus that our responses to our phenomenal experiences should ultimately be explained in terms of our normative reasons.<sup>155</sup>

It might be argued that to explain our responses to our phenomenal experiences by saying that we have normative reasons for those responses is not really to offer an informative explanation. It may not seem particularly informative to answer the question, “Why do we want to avoid agony?” with, “Because we have a reason to want to avoid agony.” But the explanation being offered is not just that we have some reason or other to want to avoid agony; it is not just that agony gives us a causal reason to want to avoid it—simply another way of saying that agony causes us to want to avoid it (the very thing to be explained). The reason we have to want to avoid agony does cause us to want to avoid it, but its content is not exhausted by this causal role: this reason not only *causes* this response, it *counts in*

---

155. Of course I am taking myself to have established that normative reasons exist (*pace* the arguments of Chapters 2 and 3). If there was no independent reason to believe that normative reasons exist, it might not make sense to explain our responses to our phenomenal experiences in terms of (our responses to our) normative reasons. But if we already know that normative reasons exist, and we know that we have some capacity to respond to these kinds of reasons, then I believe it does make sense to ultimately explain our responses to our phenomenal experiences in terms of normative reasons, because we would be explaining something that may seem mysterious (why we respond in certain specific ways to certain specific phenomenal experiences) in terms of something that we already know: that we have a capacity to respond to normative reasons.

*favour* of it (indeed it causes this response *by* counting in favour of it). And it is perfectly informative to explain why we respond to agony as we do by saying that the nature of agony counts in favour of this response (and that this is what causes this response in the first place).<sup>156</sup> The claim that the nature of agony inherently counts in favour of the aversion response tells us something that is both far from trivial—it gives us irreducibly normative information—and that we might not already know. A being who has never experienced agony, and who has never considered why those who have experienced agony respond to it with aversion, would certainly not know that this claim is true. And if we can add that there are no distorting influences acting upon us, and that the absence of such influences combined with the reason-giving nature of agony is what ultimately explains our response, then we have an explanation which is both adequately informative and

---

156. To say this is arguably to offer a *teleological* explanation for our response to agony, teleological explanations, on one interpretation, being those that

seek to explain a contingent event by showing what is *good* about that event ... The plants put out leaves because it is good for them to do so; the rain falls because it is good for it to help the plants grow; the stars move in a circular course around the earth because it is good for them to have such a perfect and beautiful motion.  
(Wedgwood 2007, 197)

Modern science completely refuses all teleological explanations for non-mental phenomena (such as the growth of plants, etc.), but this does not rule out the existence of such explanations for mental phenomena, such as our responses to our conscious experiences. For instance, we can often explain why someone forms a certain belief by pointing to the fact that it was rational for her to form this belief, and as Wedgwood points out, this appears to be a teleological explanation, because “we are explaining why a certain contingent event occurs (her forming that belief) by appeal to a certain sort of goodness—specifically, rationality—that is exemplified by that event” (197).

which does not simply replace one mystery with another: if the nature or essence of agony (its ineffable phenomenal character or *what it's like*) is inherently reason-giving, if it calls for us to respond to it in a specific way, namely, by wanting and trying to avoid it, then it is no wonder that we actually do respond to agony in precisely that way.

So the normative explanation is a reasonable explanation for the fact that we universally want to avoid agony. But that in itself does not mean that there is not a more parsimonious or consilient non-normative explanation which removes the warrant our universal response to agony appears to have. Perhaps the universal desire to avoid agony can be explained in evolutionary terms, without the need to posit any robust normative properties. The sensation of agony is generally caused by things that harm or injure us, and from an evolutionary perspective, this sensation is clearly designed to motivate us to engage in self-preserving behaviour.<sup>157</sup> The desire to avoid agony obviously has an immense survival benefit, motivating us to avoid stimuli which not only cause us agony but which threaten to harm or injure us and hence threaten our survival. On one level, therefore, it is not at all surprising that we have simply evolved to have a desire to avoid agony. But nonetheless, I don't think evolution can provide us with the complete explanation that we need. Our innate evolutionary hardwiring punishes us with agony when we engage in certain behaviours that threaten our survival, like coming directly into contact with

---

157. Figuratively speaking (I am not implying that evolution itself is in any way teleological).

dangerous stimuli, but this only presupposes our desire to avoid agony—it doesn't explain it. Nor is the survival utility of this desire sufficient to explain it in purely non-normative terms. In particular, it doesn't explain why we don't have a similar desire to avoid pleasant experiences, rather than extremely unpleasant experiences. We conceivably might have evolved to experience pleasure in response to survival-threatening stimuli, and if pleasant experiences were the kind of experiences we wanted to avoid, rather than unpleasant experiences, this desire would have the same survival utility that the desire to avoid agony actually has. But of course we haven't evolved to experience pleasure in response to survival-threatening stimuli, and we have a desire to avoid unpleasant experiences like agony rather than a desire to avoid pleasure. Evolution itself does not appear to tell us why this scenario, and not the reverse, is the one that actually obtains. Evolution only explains why we have a reward-punishment mechanism hardwired into us, rewarding us with pleasant experiences and punishing us with unpleasant experiences to motivate us to behave in a manner that is conducive to our survival. What evolution cannot explain is why pleasant experiences constitute the reward side of the reward-punishment equation and unpleasant experiences the punishment side, and thus why it is that unpleasant experiences give rise to a desire to avoid these experiences and pleasant experiences give rise to a desire to pursue these experiences, rather than the other way round.<sup>158</sup> And yet this is exactly what normative properties and normative reasons do seem able to explain.

---

158. Another way to put this point is to say that evolution only explains why

Kahane tells us that “[t]he epistemology of the badness of suffering is like the epistemology of a self-evident truth or of an immediate experience.” If an agent who has never experienced pain before were to feel pain for the very first time, when he “first feels the pain he immediately knows with certainty that this strange new experience is bad and gives reasons to get rid of it” (2010, 35). I have argued that we respond to agony by wanting and trying to avoid it *because* this phenomenal experience has the normative property of giving us a reason to want to avoid it. If I am correct that agony has this property, this does suggest that we can know that agony is bad and gives us reasons to avoid it simply by having the experience of agony, and thus without having to rely on IBE. We learned from Street (2016b) in Chapter 1 (Section 1.1) that the concept of a normative reason is acquired from the conscious experience we sometimes have of one thing’s counting in favour of or calling for another (127). And while this experience might not necessarily indicate that one thing really does count in favour of another if the thing in question is something that exists completely independently of experience, in the case of agony, there seems to be no reason to suppose that our experience of agony as inherently calling for the aversion response does not reflect the nature of agony itself.<sup>159</sup> Since,

---

we have a desire to avoid *whatever sensation it is* that accompanies injury (or the like); it does not explain why this sensation is the specific phenomenal sensation of agony.

159. Being at a red light could easily *seem* to count in favour of stopping one’s car, but not actually *be* a reason to stop one’s car. But in the case of agony, we are much closer and more intimately related as conscious beings to the phenomena which seems to give us a reason. If agony seems to count in favour of the aversion response, I believe this makes it highly probable that it actually does give us a reason

as conscious beings, we have privileged access to the inherent nature of agony, I think we can know by direct experience that the nature of agony counts in favour of the aversion response. But the point remains that the reason-giving nature of agony can also be known about on the basis of IBE.<sup>160</sup> And if we acquire new normative knowledge upon experiencing pain for the very first time, this in itself seems to entail that fundamental normative knowledge is not purely conceptual and *a priori*.

### 7.8. Conclusion

I have defended the epistemological version of normative naturalism on the grounds that fundamental normative knowledge can be acquired, not only *a priori*, but empirically by a combination of observation and inference to the best explanation. I think we can first make the observation that all conscious beings have a desire to avoid agony (at least those that actually know what agony is like) and then legitimately infer that the best explanation for this empirical fact is the normative fact that the nature of agony (it's intrinsic, ineffable phenomenal character, or *what it's like*) gives us a normative reason to want to avoid agony. The only reasonable alternative explanation might be that evolution has hardwired us to

---

for this response. We are less likely to be in error about the properties of our phenomenal experiences.

160. We might say that we can *know* that agony gives us reasons to (want to) avoid it just by having the experience of agony, but to *know that we know* that agony has the inherently reason-giving nature that it has, we have to rely on IBE. IBE indicates that our conscious experience to the effect that agony counts in favour of the aversion response is not somehow an illusion.

have this desire because of the survival benefit it provides. But this does not actually explain the thing that needs to be explained: the link between the phenomenal character of agony and our universal response to it. Only a normative fact and a normative property, like the property of being a (normative) reason, is, I believe, capable of this.<sup>161</sup> The reason why we do not find conscious beings who are attracted to experiences that they themselves find unpleasant, such as agony—not for some greater good, but entirely for their own sake—is that no being could possibly be so irrational as to desire unpleasant experiences for their own sake. And the normative fact that we all have a reason to want to avoid agony does not obtain in virtue of some more fundamental normative fact (it is not a mixed normative fact), but it can nonetheless be known *a posteriori* and is thus a case of a fundamental normative fact that is also a natural fact in the epistemological sense. Fundamental normative facts and properties, therefore, are not only deliberatively indispensable, but explanatorily indispensable as well: without them, we would not be able to explain our universal responses to certain of our experiences, such as agony, and their intrinsic phenomenal characters.

As far as a non-naturalist like Parfit is concerned,

When we consider our beliefs about the world, we often have empirical evidence for the truth of these beliefs. These beliefs can help to explain

---

161. An alternative explanation for our response to agony would only end up pushing the question back, purporting to explain this response in terms of our response to something else, and thus raising the same question about our response to this other thing. Only a normative property which is intrinsically response-guiding could answer the original question without simply raising a further question that it is unable to answer.

observable facts, and can provide testable predictions. But when we consider beliefs about what we have reasons to care about, and to do, we have no such evidence. For example, from my claim that we have reasons to want to avoid being in agony, no testable predictions follow. (2011b, 539)

Yet the Agony Principle does yield a testable prediction: the prediction that, if we know what it's like to be in agony, and we are subject to no distorting influences that prevent us from responding to agony as we have reason to respond to it, then we will, for that reason, want to avoid agony. And this prediction is no doubt confirmed: we all do want to avoid agony, and our desire to avoid agony is not the product of distorting influences—it flows directly from our knowledge of what agony is like in a way that allows no room for distorting influences to get in the way. The fact that we have this desire therefore constitutes empirical evidence for the truth of a normative belief: that we all have reasons to want to avoid being in agony. Since this is incompatible with non-naturalism, we can conclude that non-naturalism is false and that naturalism must be true. Having defended normative naturalism in this chapter, in the next chapter, we will consider whether or not normative *reductionism* is also true, that is, whether or not normative properties are *irreducibly* normative in spite of being natural.

## 8. Reductionism

Simon Blackburn (2016) defines reductionism as the thesis that “the facts or entities apparently needed to make true the statements of some area of discourse are dispensable in favour of some other facts or entities” (311). In the normative case, reductionism entails that it is not distinctly or irreducibly normative facts or entities that make normative statements true, but actually non-normative facts or entities.<sup>162</sup> For the normative reductionist, the normative is entirely constituted by and dependent on the non-normative, and normative facts and properties are therefore, in some sense, *identical* to certain non-normative facts and properties.

Since, as I have argued, there really are genuinely normative facts (Chapters 2 and 3), and there really are ontological weighty normative properties (Chapter 4), how can these facts and properties possibly be *identical* to non-normative facts and properties? Obviously normative facts and properties are not identical to facts and properties that aren’t normative. But what the reductionist should be understood to affirm is that normative facts and properties can also be specified in entirely non-

---

162. According to Blackburn, a scientific reductionist “might advocate reducing biology to chemistry, supposing that no distinctive biological facts exist, or chemistry to physics, supposing that no distinctive chemical facts exist” (311). I take a normative reductionist, therefore, to be someone who holds that there are no distinctive normative facts (without being an error theorist or a non-cognitivist, however).

normative terms.<sup>163</sup> For the reductionist, normative and non-normative claims are simply two different ways of presenting what are actually the same underlying facts and properties, facts and properties which are not themselves normative independently of being presented in a normative way. Normative facts and properties are also non-normative facts and properties, in the sense that the very same facts and properties can be presented in both normative and non-normative ways.<sup>164</sup> And for the non-reductionist, normativity is not just a way of presenting certain facts and properties; it is a way that certain facts and properties actually are in-and-of-themselves, independently of how these facts and properties might be presented by us.

The reductionist and the non-reductionist both agree that some normative claims are true and that normative facts and properties (really) exist. Streumer (2017) argues that the central difference between the two lies in what they each take to be *primarily* normative, that is, in what kind of things they each think that normativity comes from in the first place: the normative facts and properties themselves, or the kinds of things directly involved in our normative claims, like the normative

---

163. As Jackson (2016) says, “what we affirm is that ethical [or normative] properties are identical with properties that can be specified in non-ethical [or non-normative] terms” (and thus that ethical (or normative) properties can themselves also be specified in non-ethical (or non-normative) terms) (200).

164. For the reductionist, the normative claim that some act is what we ought to do could be the normative way of presenting a fact that can also be presented with the non-normative claim that this act would maximize happiness (for instance). See Chapter 6 (Section 6.3) for the distinction between sense and reference.

predicates and concepts which these claims use and the normative judgements which they express.<sup>165</sup> As he explains,

non-reductive realists [or non-reductionists] take the primary things that are normative to be properties ... They think that what makes predicates normative is that they ascribe normative properties, and that what makes judgements normative is that they apply normative predicates and thereby ascribe normative properties. So non-reductive realists think that normativity comes from properties, and extends to judgements via the predicates that ascribe these properties and express these judgements. (102)

Reductive realists (or reductionists), on the other hand, think that normativity originates from normative concepts or judgements, and then extends to the properties via the predicates and claims that are used to express these concepts and judgements.

Streumer also makes a distinction between normative properties and predicates and “descriptive” properties and predicates (1). As he defines normative and descriptive properties, a normative property is simply one that can be ascribed with a normative predicate, and a descriptive property is one that can be ascribed with a descriptive predicate (3). He doesn’t define what it is for a predicate to be normative or descriptive; he just observes that most philosophers “usually agree about which predicates are normative and which are descriptive.” Descriptive predicates are predicates such as, “is a desk,” “is white,” and “is made of wood and steel” as opposed to normative predicates like, “is right,” “is wrong,” “is good,” “is

---

165. A normative judgement is a mental item whereas a normative claim is a linguistic item. Normative claims express normative judgements and normative predicates express normative concepts and ascribe normative properties.

bad,” and “is a reason” (1).<sup>166</sup> On these definitions, normative properties can also be descriptive properties, but normative predicates cannot be descriptive predicates: a predicate is either normative or it is descriptive.<sup>167</sup> The reductionist or reductive realist holds that normative properties are identical with descriptive properties—that normative properties can also be ascribed with descriptive, non-normative predicates; the non-reductionist or non-reductive realist holds that normative properties are *not* identical to descriptive properties—that normative properties can only be ascribed with normative predicates and never descriptive ones.

This explains why the reductive realist cannot take the primary things that are normative to be properties. As Streumer says, “if reductive realism is true, the difference between normative and descriptive properties is a difference in language that is not matched by a difference in the nature of these properties” (103). If normative properties are identical to certain descriptive properties, there is of course no difference in the nature of these properties; normative properties would also be descriptive properties in this scenario. There would only be a difference in the predicates—normative and descriptive or non-normative—that can be used to ascribe these properties. All properties, normative or otherwise, would be descriptive in the sense that they can be ascribed using descriptive or non-normative predicates. But a subset of these properties would also be ascribable using normative

---

166. Streumer adds that “is a reason” is a normative predicate only if it is equivalent to “counts in favour.”

167. Streumer doesn’t explicitly state this, but I think it is implied by the examples of descriptive predicates that he uses.

predicates, and this is what would make these properties normative: the fact that they can be ascribed using normative (as well as non-normative) predicates.

Suppose, for instance, that the normative property of being what we ought to do is identical to the descriptive property of maximizing happiness. There would then be a single property that the predicates, “being what we ought to do” and “maximizing happiness” both ascribe, and this property would be both normative and descriptive, because it can be ascribed both by a normative predicate (“being what we ought to do”) and a descriptive predicate (“maximizing happiness”). Since this single property can be ascribed using both normative and non-normative predicates, this property’s normativity cannot originate from the property itself and extend to the predicates that ascribe it, because this would make the non-normative predicate (“maximizing happiness”) into a normative predicate as well. Instead, the normativity must extend from the normative predicate (“being what we ought to do”) to the property, through the fact that this property can be ascribed using this normative predicate. This predicate’s normativity probably comes not from the predicate itself, but from the normative concepts or judgements that it can be used to express, but if normative properties can also be ascribed with non-normative predicates, as the reductionist holds, it follows that the normativity of normative predicates is prior to the normativity of normative properties. Whereas for the non-reductionist, of course, the normativity of normative properties is prior to the normativity of normative predicates.

In what follows, I will first outline the reductionist's case for the conclusion that normative properties are identical to descriptive properties (Section 8.1), based as it is on the idea that normative and descriptive properties are necessarily co-extensive and that this provides a strong reason to identify them (Jackson 2016, 202). I will then provide a critique of the reductionist's position in Section 8.2 (a critique that applies to a version of reductionism which I believe avoids Streumer's own objections). In Section 8.3, I will concede the reductionist's claim that obviously normative predicates like "is right" do in fact ascribe the same properties as certain disjunctive predicates which are not obviously normative, such as the predicate Streumer refers to as predicate D\*. But I will argue, contrary to the reductionist, that predicates like D\* are actually normative and not descriptive predicates: predicate D\* does not ascribe a descriptive property, but a *pattern* in descriptive properties which is not itself descriptive but is in fact normative, since it is identical to the normative property of being right. And finally, in Section 8.4, I will argue that the patterns in descriptive properties ascribed by predicates like D\* are not only normative but irreducibly normative.

### 8.1. *The Reduction Argument*

Jackson and Streumer both defend a similar argument against non-reductive normative realism, an argument which Streumer calls "the Reduction Argument" (2017, 9). The Reduction Argument states that:

- (1) Normative predicates are necessarily co-extensive with descriptive predicates.<sup>168</sup>
- (2) Two predicates ascribe the same property if and only if they are necessarily co-extensive.<sup>169</sup> (11)
- (3) Therefore, normative properties (if they exist) are identical to descriptive properties.<sup>170</sup> (40)

It this argument is sound, it follows that normative properties can also be ascribed with descriptive, non-normative predicates, and thus that the normativity of predicates is prior to the normativity of properties, and that normative concepts or normative judgements are the things that are primarily normative, rather than the normative properties themselves.

Since normative properties are *grounded* in descriptive properties—things have the normative properties that they have in virtue of also having certain descriptive properties—normative properties (and predicates) are necessarily co-extensive with descriptive properties (and predicates), so long as normative properties also *supervene* on descriptive properties. Everything that has normative properties also has descriptive properties (9); there is obviously nothing which has only normative properties and no descriptive properties. And if it is also the case that there can be no difference in an object's normative properties without a corresponding difference in its descriptive properties (Jackson 2016, 197), it follows

---

168. Or as Jackson puts it, “ethical [or normative] properties are necessarily co-extensive with non-ethical [or non-normative] properties” (2016, 197).

169. This is the criterion of property identity which Streumer labels “(N)”.

170. Normative properties of course don't actually exist according to Streumer's error theory.

that, whenever an object has certain normative properties, it must also have certain descriptive properties, which in turn implies that normative and descriptive properties (and predicates) are necessarily co-extensive.

But why think that normative properties supervene on descriptive properties in the first place? Streumer takes it to be a “central thought” about normative properties that “[f]or all possible worlds  $W$  and  $W^*$ , if the instantiation of descriptive properties in  $W$  and  $W^*$  is exactly the same, then the instantiation of normative properties in  $W$  and  $W^*$  is also exactly the same” (2017, 6).<sup>171</sup> Jackson simply calls Supervenience “plausible” (2016, 196) without providing any explicit defence of it. And Enoch (2013) admits that “he doesn’t know of any compelling arguments with this supervenience claim as its conclusion.” But Enoch also tells us that “the thought that a mere difference in modal location—in which possible world something is—can all alone, without a natural difference, make (or indicate) a normative difference seems utterly implausible to me, barely intelligible even” (141). Ernest Sosa (1980) likewise affirms that

supervenience ... characterizes normative and evaluative properties generally. Thus, if a car is a good car, then any physical replica of that car must be just as good. If it is a good car in virtue of such properties as being economical, little prone to break down, etc., then surely any exact replica would share all such properties and would thus be equally good. (15)

I think we should accept that Supervenience is indeed a conceptual truth about the nature of normative properties. We saw in the previous chapter that fundamental normative knowledge can be acquired on the basis of empirical evidence alone

---

171. This thought is meant to reflect the nature of normative properties.

(without having to rely on *a priori* normative assumptions). And this would surely be impossible if normative properties didn't supervene on descriptive properties.

So normative properties certainly seem to be necessarily co-extensive with descriptive properties, and premise (1) of the Reduction Argument therefore seems to be true. But does this really mean that normative properties are identical to descriptive properties—that normative properties can also be ascribed with non-normative predicates, and thus that properties are not the original source of normativity? Enoch denies that necessarily co-extensive properties are necessarily identical, arguing that whatever reason there is to think that such properties are identical basically comes down to considerations of parsimony—in particular, the thought that “multiplying distinct necessarily co-extensive properties does not add explanatory power” (2013, 139). Enoch’s response to this thought is that, although normative properties may not add explanatory power beyond that of the descriptive properties with which they are necessarily co-extensive, normative properties still “satisfy the parsimony requirement by being deliberately indispensable” (140) and therefore don’t need to be identified with descriptive properties for the sake of preserving parsimony.

We can agree that normative properties are deliberately indispensable: as we saw in Chapter 3, the likes of Occam’s Razor and IBE cannot be used to refute the ontological commitments of normative deliberation. But this does not mean that there is no gain in parsimony involved in identifying normative properties with descriptive properties. Reductive realism is more ontologically parsimonious than

non-reductive realism (even naturalistic non-reductive realism) because it does not posit distinct necessarily co-extensive properties. All else being equal, therefore, reductive realism will come out ahead on plausibility points for this reason.

Streumer points out that, even if normative properties are deliberately indispensable, this does not mean that *irreducibly* normative properties are deliberately indispensable (2017, 23). And we have already rejected Enoch's and Parfit's arguments (based as they are on the just-too-different-intuition) for the conclusion that normative facts and properties must be non-natural and metaphysically irreducible (Chapter 6).

Unlike Enoch, however, we should also reject the thought that normative properties add no explanatory power of their own beyond that of non-normative properties.<sup>172</sup> By the argument of the previous chapter, only normative properties could adequately explain our universal responses to certain of our phenomenal experiences, such as agony: an explanation that appeals entirely to facts and properties that are in no way normative (I have argued) would always fail to explain the link between the specific phenomenal character of the experience in question and the specific response to which it universally gives rise. But just because normative properties (such as the property of being a reason for aversion, or intrinsic *ought-to-be-avoidedness*) are needed to explain these kinds of responses does not mean that these properties are not also descriptive. A descriptive property is not necessarily a

---

172. That is, properties that aren't normative.

property that isn't normative, and to assume that descriptive properties cannot be normative would already be to beg the question against reductive realism.

Suppose that the normative property of being a reason for the kind of intense, instantaneous aversion that always follows the experience of agony is necessarily co-extensive with the (apparently descriptive) property of being extremely unpleasant (since the phenomenal character of agony necessarily instantiates both of these properties).<sup>173</sup> And suppose that premise (2) of the Reduction Argument is true and that necessarily co-extensive properties are necessarily identical. In that case, the descriptive property of being extremely unpleasant would also be a normative property, and what explains why we respond to agony (given that it has this property) with this kind of aversion would not be an irreducibly normative property, but the fact that a descriptive property (being extremely unpleasant) is also a normative property (being a reason for intense, instantaneous aversion). Is there any gain in explanatory power involved in positing an *irreducibly* normative property to explain our response to agony? If the phenomenal characters of experiences which have the property of being extremely unpleasant also have the irreducibly normative property of being a reason for intense, instantaneous aversion, this would indeed explain why we respond with such aversion to those experiences. But this would also seem to be explained equally well if the property of being extremely unpleasant is itself identical to the property of being a reason for this kind

---

173. I say "suppose" because, although anything that has the property of being extremely unpleasant (in the phenomenal sense) also has the property of being a reason for this kind of aversion, the reverse might not follow.

of aversion. Such an explanation would tell us that we respond in a certain way to a certain descriptive property because that property is also a normative property. And the reductive explanation, which simply says that a well-known descriptive property has the second-order property of being normative, is definitely more parsimonious than the non-reductive explanation, which says that this descriptive property is necessarily co-extensive with, but not identical to, an irreducibly normative property.<sup>174</sup>

I think this illustrates that the burden on proof is on the non-reductionist to show that necessarily co-extensive properties are not identical, in light of the fact that it is more parsimonious to identify these properties than to regard them as distinct. The necessary co-extension of normative and descriptive properties does provide a strong (though not apodictic) reason to identify them (Jackson 2016, 202). The typical strategy for the non-reductionist is to attempt to offer various examples of necessarily co-extensive properties which are nonetheless distinct. But Streumer argues that if properties are what he calls *ways objects can be*—meaning that “what it is for an object to have a certain property is that this object itself is a certain way” (2017, 12)—it follows that (N), according to which necessarily co-extensive properties are identical, is the correct criterion of property identity. He considers

---

174. In other words, the reductive explanation says that we respond to agony in the way that we do because the (descriptive) property that agony has of being extremely unpleasant is itself a normative property, whereas the non-reductive explanation says that we respond to agony in this way because, in addition to having the property of being extremely unpleasant, it also has the (irreducibly normative) property of giving us a reason to respond in that way. It is clear that the former explanation is more parsimonious than the latter.

several purported counterexamples to (N), arguing that “if properties are ways objects can be, what (N) says about these examples is exactly right” (13). Probably the most plausible of these alleged counterexamples is the claim that, “[t]he predicate ‘is the only even prime number’ and ‘is the positive square root of four’ are necessarily co-extensive but ascribe different properties” (16). Parfit (2017), as a non-reductive realist, claims that “it’s one thing to be the positive square root of 4, and quite a different thing to be the only even prime number” (69). And Streumer responds that, since the phrases, “the positive square root of four” and “the only even prime number” both refer to the number two (as Parfit admits), the predicates, “is the positive square root of four” and “is the only even prime number” must ascribe a single property: the property of being the number two (2017, 17).

The predicates, “is the positive square root of four” and “is the only even prime number” no doubt express different concepts, and therefore ascribe different properties if properties are what Streumer calls *shadows of concepts*, which means that “what it is for an object to have a certain property is that this object falls under a certain concept” (12). Parfit tells us that he is using the word “property” in what he calls the *pleonastic* sense, the sense “in which any claim about something can be restated as a claim about this thing’s properties,” and thus in which the use of the word “property” adds nothing to the content of our claims.<sup>175</sup> He also calls this sense the *description-fitting* sense, because in this sense, properties “fit the descriptive

---

175. For instance, as he puts it, “[i]nstead of saying that the Sun is bright, or that some argument is valid, we can say that the Sun has the property of being bright, or that this argument has the property of being valid” (66).

words or phrases with which we refer to them" (2017, 66) and he contrasts this with the *necessarily co-extensional* sense, in which (presumably) it is analytic that necessarily co-extensive properties are identical (or that necessarily co-extensive predicates ascribe the same property) (68). He then claims that "*being the only even prime number* and *being the positive square root of 4* are the same property in the necessarily co-extensional sense," but that "these properties are different in the description-fitting sense." And why is this? Because "[t]he concept of *being the only even prime number* does not describe, and thereby refer to, the property of *being the positive square root of 4*" (69).

The above seems to indicate that, as far as Parfit is concerned, the properties of being the only even prime number and being the positive square root of four are different only because the relevant *concepts* are different. Streumer would agree that these properties are different in the description-fitting sense (which seems to be the same thing as taking properties to be shadows of concepts). And they would both also agree that these properties are the same in the necessarily co-extensional sense (because this sense simply repeats the fact that these properties are necessarily co-extensive). So this example doesn't show that these necessarily co-extensive properties are nonetheless distinct if we take properties not to be shadows of descriptions or concepts but ways objects can be in-and-of-themselves.

Parfit considers a further sense of "property" when he says that this word "is often used, not in this description-fitting sense [or the necessarily co-extensional sense], but in some narrower, ontologically weighty sense" (68). And as someone

who denies that normative truths have ontological implications, he doesn't think that this sense applies to normative properties (indeed there are no normative properties, according to Parfit, in this ontological sense).<sup>176</sup> But as we saw in Chapter 4, normative properties do have ontological implications, and so this ontologically weighty sense of "property" applies to normative properties just as well as the description-fitting and necessarily co-extensional senses. It is basically incontestable that, if two concepts are different, the properties corresponding to these concepts are also different *in the description-fitting sense* (or the sense in which properties are shadows of concepts), and also that, if two properties are necessarily co-extensive, they are the same property *in the necessarily co-extensional sense*. And so it is this ontological sense that is the relevant sense when we are asking whether or not normative properties are identical to descriptive properties. Since (as we also saw in Chapter 4) what it is for a property to have ontological implications is for it to have some kind of causal standing or efficacy, we can say that properties are identical in this particular sense if and only if they necessarily have the exact same causal powers and effects.<sup>177</sup>

---

176. There are no ontologically weighty normative properties according to Streumer as well of course (because there are no normative properties).

177. The ontological sense does not seem to be the relevant sense for properties like being the only even prime number or being the positive square root of four. There is no possible counterfactual scenario in which the only thing that actually has these properties (the number two) does not have these properties, meaning that these properties do not support counterfactuals and thus (under the counterfactual theory of causation at least) do not have any kind of causal standing, nor any ontological implications as a result.

Is it possible for properties to be necessarily co-extensive yet have different causal powers or effects? I think the answer must clearly be no. If properties A and B are necessarily co-extensive, there is no possible counterfactual scenario in which an object has property A without also having property B. This means that there is no possible counterfactual scenario in which an object with property A has different causal powers or effects than it actually has because it doesn't have property B (and *vice versa*). We therefore cannot say truly that an object which has properties A and B would have different causal powers or effects than it actually has if it had property A but not property B (or *vice versa*). Neither property A nor property B, considered as distinct properties, makes a difference to the causal powers or effects of any object which has these properties. And if objects with property A necessarily have exactly the same causal powers and effects as objects with property B, it surely follows that property A itself has exactly the same causal powers and effects as property B, and thus that A and B are the same property, at least in the ontological sense.

So (N) is the correct criterion of property identity for properties that are causally efficacious, ontologically weighty entities in their own right: if such properties are necessarily co-extensive, they are in fact the same property. Necessarily co-extensive *concepts* are not necessarily identical, but this is not something that is at issue between the reductionist and the non-reductionist. Robust normative properties are not merely shadows of concepts (or descriptions); they have the same ontological implications as the descriptive properties with which they are necessarily co-extensive. It follows from this that (N) is the correct criterion of

property identity for normative properties. This seems to give us premises (1) and (2) of the Reduction Argument, which together entail the conclusion that normative properties are identical with descriptive properties, and thus that normativity originates from concepts or judgements rather than the normative properties themselves.

### 8.2. *The problem with reductionism*

The Reduction Argument seems to establish that normative properties are identical to descriptive properties. If normative properties are identical to descriptive properties, normative predicates do not ascribe irreducibly normative properties, but rather descriptive properties. And as Streumer points out, “[i]f a certain normative predicate ascribes a certain descriptive property, this cannot be a brute fact. There must be something that makes it the case that *this* normative predicate ascribes *this* descriptive property” (2017, 43). If the normative predicate, “is right” ascribes the descriptive property of maximizing happiness, there must be some reason why this predicate ascribes the property of maximizing happiness, rather than, say, the property of maximizing suffering, and some reason why it is not some other, conflicting normative predicate (such as “is wrong”) that ascribes this property instead. The reductionist needs to be able to explain why normative predicates ascribe the particular descriptive properties that they ascribe.

Predicates, of course, express concepts, so if a certain normative predicate ascribes a certain descriptive property, this must be in virtue of the concepts

expressed by both the normative and descriptive predicates that ascribe the same property. While the concepts expressed by the predicates, “is right” and “maximizes happiness” are different (one is normative, the other descriptive), if these predicates ascribe the same property, it must be in virtue of a necessary relation between these two concepts that these predicates ascribe the same property, and thus that the normative predicate ascribes the particular descriptive property that it ascribes. If something cannot fall under the concept expressed by the predicate, “is right” without also falling under the concept expressed by the predicate, “maximizes happiness,” because of the necessary relation standing between these two concepts, then this would explain why the normative predicate, “is right” ascribes the descriptive property of maximizing happiness. In other words, it is conceptual normative truths (such as, perhaps, the truth that what is right is maximizing happiness) which explain why certain normative predicates ascribe certain descriptive properties.

However, if it is conceptual normative truths—and hence the necessary relations between normative and descriptive concepts—that explain why certain normative predicates ascribe certain descriptive properties, in virtue of what do these conceptual truths—and these conceptual relations—hold? Concepts (on the traditional view at least) “are abstract, sharable, mind-independent ways of thinking about objects and their properties” (Cuneo and Shafer-Landau 2014, 409), and “they are also *referential devices* or *ways of getting things in mind* that enable thinkers to refer to things such as objects and properties.” Concepts have *essences* in virtue of which

they apply only to certain things and not others: “[i]t belongs to the essence of ‘being wrong,’ for example, that it applies to exactly those things that are wrong” (410). If “is right” ascribes the property of maximizing happiness, then it must belong to the essence of the concept expressed by this normative predicate that it applies exactly to those things that have the descriptive property of maximizing happiness. And it must also belong to the essence of the concept expressed by the descriptive predicate, “maximizes happiness” that it applies exactly to those things that have the normative property of being right (thus generating the conceptual truth that what is right is maximizing happiness).<sup>178</sup>

But can it just be a brute fact that it belongs to the essence of a certain normative concept that it applies to exactly those things that have a certain descriptive property? Take a non-normative conceptual truth such as the truth that nothing can be red and green all over. The essences of the concepts expressed by the predicates, “is red all over” and “is green all over” are such that nothing that satisfies the first can possibly satisfy the second. And this does not appear to be a brute fact: the essences of the concepts are necessarily related in the specific way that they are because the *properties* that these concepts represent—the properties of being red all over and of being green all over—are necessarily related in the same way. Nothing can have the *property* of being red all over at the same time as having the

---

178. As Cuneo and Shafer-Landau say, the “proposition that x is F is a conceptual truth if it belongs to the essence of ‘F’ that, necessarily, anything that satisfies ‘x’ also satisfies ‘F’” (410). And this is what would be the case for the proposition that what is right is maximizing happiness if the predicate, “is right” ascribes the property of maximizing happiness.

*property* of being green all over, and this is what explains why nothing can satisfy both of the equivalent concepts at the same time. Concepts being ways of thinking about objects and properties, if properties are ways objects can be, the essences of concepts and the relations between them are grounded in the essences of the underlying properties: it is impossible to *think* of something being both red and green all over at the same time because it is impossible for something to *be* both red and green all over at the same time. And the same should apply in the normative case. If the essences of the *concepts* expressed by the predicates, “is right” and “maximizes happiness” are such that nothing can satisfy the one without also satisfying the other, this must be explained by the fact that the essences of the equivalent *properties* are such that nothing can have one property without also having the other.

If normative properties really are identical to descriptive properties, it must be the relations between normative and descriptive concepts that explain why certain normative predicates ascribe certain descriptive properties (and why certain descriptive predicates ascribe certain normative properties), in virtue of the fact that normative and descriptive predicates express these particular concepts.<sup>179</sup> And since

---

179. Streumer says that, under reductive realism, what makes it the case that a certain normative predicate ascribes a certain descriptive property is either that, in certain *descriptively* specified conditions, users of this predicate would apply it to objects that have this property, or that, in certain *normatively* specified conditions, users of this predicate would apply it to these same objects (2017, 43). He then argues that, in the first case, reductive realism faces what he calls *the false guarantee objection*: it falsely implies that there are certain descriptively specified conditions in which people’s normative judgements are guaranteed to be correct (he takes it to be

concepts are ways of thinking about objects and properties, it must be the relations between objects and properties themselves that explain the relations between concepts: in other words, an object or a property can be thought about in a certain way only because the object or property can itself actually be that way. Under normative reductionism, the relations between normative predicates and descriptive properties (and descriptive predicates and normative properties) are explained by certain normative conceptual truths, and so it must be the properties represented by the concepts involved in these truths that explain the relations between normative and descriptive concepts, and hence why certain normative predicates ascribe certain descriptive properties. But if this is indeed the case, it must be the normative properties that are primarily normative, rather than the normative concepts: if the relations between normative and descriptive concepts are explained by the relations between normative and descriptive properties, the normativity of the normative properties must be prior to the normativity of normative predicates and concepts.

And this is entirely contrary to the hypothesis that normative properties are identical to descriptive properties, and thus to the hypothesis of normative reductionism.

---

a central thought about normative judgments that there are no descriptively specified conditions in which people's normative judgements are guaranteed to be correct) (53). And in the second case, reductive realism faces what he calls the *regress objection*: it implies that a normative judgement, applying a certain normative predicate to an object with a certain descriptive property, would only be true if a further normative judgement, according to which the normatively specified conditions (in which this normative predicate ascribes that descriptive property) actually obtain, is also true, and so on *ad infinitum*, generating a vicious regress (57). I think the view that normative predicates ascribe the same properties as certain descriptive predicates because of the necessary relations between the concepts these predicates express avoids both of these objections.

So it appears that, *pace* the Reduction Argument, since normative properties are necessarily co-extensive with descriptive properties, they must be identical to descriptive properties, in the sense that normative predicates ascribe properties that can also be ascribed with descriptive predicates. But it also appears that, since normative and descriptive predicates ascribe the same properties in virtue of the relations between normative and descriptive concepts, and relations between concepts obtain in virtue of the relations between properties, it is the normative properties that are primarily normative rather than the concepts. And as we have seen, if normativity originally comes from properties, any predicate that ascribes a normative property will itself be a normative predicate in virtue of ascribing that property, meaning that no predicate that ascribes a normative property will be a descriptive predicate.<sup>180</sup> Since descriptive predicates do not ascribe normative properties, therefore, normative properties are not actually identical to descriptive properties, and normative predicates do not actually ascribe descriptive properties, even though normative properties and predicates are necessarily co-extensive with descriptive properties and predicates. How can this be so?<sup>181</sup>

---

180. Remember, that while normative properties can also be descriptive properties (at least in so far as the definitions of “normative property” and “descriptive property” are concerned), normative predicates (by definition) cannot also be descriptive predicates.

181. Streumer would of course answer that this is because there are in fact no normative properties. But quite aside from the fact that I have already defended metanormative realism (Chapters 2 and 3), I don’t think that we have to take this radical line (for reasons I am about to explain).

### 8.3. Predicate $D^*$

Jackson and Streumer both argue that the normative predicate, “is right” is necessarily co-extensive with a descriptive predicate which Jackson refers to as  $N$  (2016, 201) and Streumer refers to as  $D^*$  (2017, 10).<sup>182</sup> Predicate  $D^*$  is a disjunctive predicate that consists of all the descriptive predicates that are satisfied by anything, in all possible worlds, that also satisfies the normative predicate, “is right” (9-10). In other words,  $D^*$  is a disjunction of all the descriptive predicates that exist *apart* from those that are specifically excluded by the normative predicate, “is right”: every descriptive predicate that anything could possibly satisfy so long as it also satisfies the predicate, “is right.”<sup>183</sup> And if something which satisfies the predicate, “is right” also satisfies any of the descriptive predicates in this disjunction (which it of course does), it also thereby satisfies predicate  $D^*$ , from which it follows that “is right” is necessarily co-extensive with  $D^*$ . Given the criterion of property identity ( $N$ ) appealed to by the Reduction Argument (and which I have already defended), “is right” ascribes the same property as predicate  $D^*$ .

According to Jackson, “one thing we cannot say ... is that ethical [or normative] properties match up with pattern-less infinite disjunctions of natural [or descriptive] properties” (2016, 201). Streumer points out that, “if properties are ways

---

182. I will refer it as  $D^*$  to avoid confusion with ( $N$ ).

183. Satisfaction of the predicate, “is right” naturally excludes the satisfaction of predicates such as, “involves engaging in the recreational slaughter of others” and the like—any predicate the satisfaction of which by something prevents that thing from being right (and thus from satisfying the predicate, “is right”).

objects can be [they are not merely shadows of descriptions or concepts], properties are not themselves disjunctive or non-disjunctive"; they can only be ascribed with disjunctive or non-disjunctive predicates (2017, 26). So while the *predicate*, "is right" ascribes the same property as predicate D\*, an infinitely disjunctive predicate consisting entirely of descriptive predicates, the *property* of being right cannot be equivalent to an infinite disjunction of the descriptive properties ascribed by the disjuncts of predicate D\*. The property of being right must instead be equivalent, not to these descriptive properties themselves, but to a *pattern* in these descriptive properties. And it is this pattern that makes the property normative, not the fact that it can be ascribed by a normative predicate. What this means is that the property ascribed both by "is right" and predicate D\* is not only normative but primarily normative. Normativity originally comes not from predicates, concepts or judgements, but from certain patterns in descriptive properties; and these patterns in descriptive properties are what normative properties are. Since the property ascribed by predicate D\* is primarily normative, it therefore cannot be descriptive (because any predicate which ascribes this property will thereby be normative, not descriptive). And so it is false that the normative property of being right is identical to a descriptive property.

This implies that premise (1) of the Reduction Argument is actually false: normative predicates and properties are not actually necessarily co-extensive with descriptive predicates and properties. Normative properties are necessarily co-extensive with patterns in descriptive properties, and since these patterns in

descriptive properties are the things that are primarily normative, and these patterns are equivalent to normative properties like the property of being right, predicates like  $D^*$  which ascribe these properties are not actually descriptive predicates (because they ascribe properties that are not only normative but primarily normative). As such, it is in fact not the case that normative properties can also be ascribed by descriptive predicates (or that normative predicates also ascribe descriptive properties).

Streumer argues that predicate  $D^*$  must be a descriptive predicate: in his words, "since a predicate that wholly consists of descriptive predicates is itself a descriptive predicate, predicate  $D^*$  is a descriptive predicate" (10). But I think this assumption commits the fallacy of composition: we cannot assume that, just because the individual parts of predicate  $D^*$  are descriptive, the whole predicate itself must be descriptive. If predicate  $D^*$  ascribes a property that is also ascribed by the predicate, "is right," then it ascribes a normative property. The property ascribed by both these predicates is a pattern in properties which consists entirely of descriptive parts. But this does not prevent this pattern as a whole from being a normative property. So the fact that predicate  $D^*$  consists entirely of descriptive parts should not prevent it from being a normative predicate either.<sup>184</sup> If normative predicates are

---

184. Especially when we remember that Streumer doesn't tell us what normative and descriptive predicates actually are. If we start by defining normative and descriptive properties as those which can be ascribed with normative and descriptive predicates (respectively), and do not define what it is for a predicate to be normative or descriptive, we are quite likely to end up with the conclusion that the normativity of predicates is prior to the normativity of properties, because we

simply those that ascribe normative properties—since normativity extends from properties to the predicates that ascribe them—this will make predicate  $D^*$  into a normative predicate. And I think we should conclude from the Reduction Argument, not only that the normative property of being right is identical to the (normative) property ascribed by predicate  $D^*$ , but that predicate  $D^*$  is itself normative because it ascribes the normative property of being right, and because this property is identical, not to any descriptive property, but to a normative pattern of descriptive properties and is therefore primarily normative. Since predicate  $D^*$  is a normative predicate if it ascribes a property that is primarily normative, and the property of being right is, I submit, just such a property, predicate  $D^*$  is in fact a normative predicate. Thus it is not in fact the case that the normative predicate, “right” is necessarily co-extensive with a descriptive predicate.

But how can premise (1) of the Reduction Argument be false if Grounding and Supervenience are both clearly true? Normative properties *do* supervene on descriptive properties, and things *do* have whatever normative properties they have

---

are defining the normativity of properties in terms of the normativity of predicates. In that case, it will be difficult to see how a predicate like  $D^*$  could possibly be normative. But if instead we define a normative predicate as one which ascribes a normative property (while leaving it open exactly what a normative property is), we can easily see how  $D^*$  could be a normative predicate even though it consists entirely of descriptive parts: the parts of this predicate only ascribe parts of the pattern which constitutes the property of being right, parts which are not themselves normative, whereas the whole predicate ascribes every part of the pattern—a pattern which (being equivalent to the property of being right) certainly is normative. The whole predicate is normative, even though its parts are not, because together its disjuncts ascribe every part of a pattern which is normative, even though the individual parts of the pattern are not.

in virtue of also having certain descriptive properties, so how can normative properties fail to be necessarily co-extensive with descriptive properties? I believe the answer is that, although any normative property is necessarily co-extensive with *some descriptive property or other*, no normative property is necessarily co-extensive with *any descriptive property in particular*. Normative properties are necessarily co-extensive, not with any particular descriptive properties, but with patterns in descriptive properties that can be ascribed by predicates like  $D^*$ , and this is why normative properties are equivalent to these patterns of descriptive properties, and thus why they can be ascribed by predicates like  $D^*$  (which, I have argued, are in fact normative predicates). So premise (1) is not so much false as ambiguous: it could mean either that normative properties and predicates are necessarily co-extensive with *particular* descriptive properties and predicates, or that they are necessarily co-extensive with *some descriptive properties and predicates or other*. In the first case, the conclusion—that normative properties are identical with descriptive properties, and that normative and descriptive predicates ascribe the same properties—will follow, but the premise is false. In the second, premise (1) is indeed true, but the conclusion will not follow: it is not obviously descriptive predicates (such as “maximizes happiness”) which ascribe the same properties as obviously normative predicates (such as “is right”), but predicates like  $D^*$ , which may not be obviously normative, but which are not obviously descriptive either, and which must in fact be normative because they ascribe properties that are primarily normative.

The Reduction Argument, therefore, according to which normative properties (if they even exist at all) are identical to properties that can also be ascribed using descriptive, non-normative predicates, fails. We can construct a sound argument for the conclusion that normative properties are identical, not to descriptive or non-normative properties, but to patterns in such descriptive or non-normative properties (patterns which are not identical to the descriptive or non-normative properties themselves):

- (1\*) Normative properties are necessarily co-extensive with patterns in descriptive properties (like that ascribed by predicate D\*).
- (2\*) Properties are identical if and only if they are necessarily co-extensive (N).
- (3\*) Therefore, normative properties are identical to patterns in descriptive properties.

I think this argument establishes that properties are the things that are primarily normative, rather than normative predicates, concepts or judgements. The normativity of normative properties is prior to the normativity of normative predicates. Normative properties do not get their normativity from being ascribed by normative predicates; normative predicates get their normativity from ascribing normative properties.

#### 8.4. *Is this reductionism?*

The non-reductionist is correct to hold that normative properties are the things that are primarily normative. But nonetheless, the above argument shows that there would be no normative properties whatsoever if not for certain non-normative

properties and the patterns these properties form<sup>185</sup>: normative properties, while perfectly real, are not *fundamental* features of the universe. Doesn't this imply that there are no irreducibly normative properties, and thus that non-reductive metanormative realism is false?

Jackson would no doubt maintain that to identify normative properties with patterns in non-normative properties is exactly to provide a reductive account of normativity. Drawing an analogy with the physical property of *density* (which is identical to mass divided by volume) he argues that “we identify density with a pattern of properties that aren't density, namely, the ratio of mass to volume,” and “[i]t is in virtue of this fact that we can think of the identification of density as a *reductive* account” (2016, 199). But of course, to identify density with a pattern of properties that aren't density (that is, the *ratio* of mass to volume) is not to identify density with the non-density properties (mass and volume) themselves. Density itself is something over-and-above the non-density properties on which it nonetheless ontologically depends. It is identical neither to mass or volume individually, nor just to the combination of mass and volume, but to a specific relationship between these two properties: mass *divided by* volume. An object which has mass, volume and density has three distinct properties, none of which is identical or reducible to any other. And the ratio of an object's mass to its volume is neither identical nor reduces to a property that isn't density—it *is* density.

---

185. Or perhaps, no normative property would be instantiated if not for certain non-normative properties being instantiated (if all properties exist necessarily).

Lee Smolin (2015) describes properties like density, which are always properties of a composite system “that would not make sense when applied to the parts [of the system],” and which “are not of the same kind as those of its component parts or processes,” as *emergent* properties. Emergent properties must be added to the list of properties in addition to the properties of the component parts of the system (378). In his illustration, “a litre of gasoline can have a mass, a momentum, a temperature, and a density. Its component molecules have mass and momentum, but it makes no sense to talk about the temperature or density of a molecule ... So we say that temperature and density are emergent properties of the gas” (378-79). Smolin claims that “[t]his common circumstance does not represent a limit of the method of reductionism; instead it represents its intensification. For emergent properties can often be elucidated in terms of the properties of the parts” (379). But this is because he defines reductionism simply as “the good advice that if you want to understand a composite system, which means a system composed of parts, you will do well by explaining the properties of the composite system in terms of the properties of the parts” (378). As naturalists, we should be trying to explain the normative properties of things in terms of their non-normative properties, but this just means explaining how their non-normative properties *give rise* to their normative properties; it does not mean that normative properties are in any way identical to non-normative properties, any more than density is identical to the properties of mass or volume. If my arguments are correct, normative properties are

emergent properties which can be explained in terms of more basic non-normative properties, but this does not compromise their irreducible normativity.

Returning to Blackburn's definition of reductionism with which this chapter started (the facts or entities apparently needed to make true the statements of some area of discourse are dispensable in favour of some other facts or entities (2016, 311)), we can observe that statements about an object's density are made true by facts about the object's mass and volume, *plus* the fact that density is mass divided by volume. Given this fact about what density is, the fact that an object has a certain density is of course equivalent to the fact that it has a certain mass and volume. But neither of these facts is the same fact as the fact that density is mass divided by volume. And if not for this further fact, objects would not have density, and statements about the densities of objects would not be true. In the normative case, since being right is equivalent to  $D^*$ , the fact that something is right is equivalent to the fact that it has the property ascribed by predicate  $D^*$ . But these facts are not equivalent to the fact that being right is  $D^*$ . And if not for this further fact, nothing would be right, and statements about which things are right would not be true.

So what, then, is the status of the fact that being right is equivalent to  $D^*$ ? Is it an irreducibly normative fact? If it isn't, I think it would follow that irreducibly normative facts are not needed to make normative statements true, and that we can therefore say that reductionism, as a general thesis, applies to the normative domain. The fact that something is right (or that it is  $D^*$ ) may be a normative fact, but it is not an irreducibly normative fact unless the fact that being right is  $D^*$  is itself an

irreducibly normative fact. If something is right, what makes it the case that this fact obtains is simply that the thing has certain descriptive properties sufficient for it to satisfy predicate  $D^*$ , combined with the fact that being right is  $D^*$ . And since the fact that being right is  $D^*$  involves a certain normative property being identical to a certain normative property ("being right" and " $D^*$ " of course ascribing the very same normative property), it might seem that it could not possibly be an irreducibly normative fact. But as we have seen, what this fact actually means is that when non-normative properties come together in a certain way, a normative property is thereby formed. The fact that being right is  $D^*$  tells us thus that, when something has certain non-normative properties (sufficient for it to be  $D^*$ ), it will also have a certain normative property (being right).<sup>186</sup> This and other facts like it are indispensable to the truth of statements within the normative domain. I believe, therefore, that this fact *is* an irreducibly normative fact. And so is any fact according to which a certain pattern in non-normative properties equates to a certain normative property.

### 8.5. Conclusion

I conclude that the normative cannot be completely reduced to the non-normative, and thus that normative reductionism is false. This is in spite of the fact that an argument close to the Reduction Argument actually succeeds. But as that argument shows, normative properties are not identical to straightforwardly

---

186. Just like the fact the density is mass divided by volume tells us that, if an object has the properties of mass and volume, it will also have the property of density.

descriptive properties (such as maximizing happiness), but to patterns in descriptive or non-normative properties, patterns which are not reducible to the non-normative properties themselves. Normativity originally comes from these patterns in non-normative properties, not from the predicates that ascribe the properties to which these patterns are equivalent, nor the concepts or judgements which these predicates express. And since normative properties are equivalent to patterns in non-normative properties, rather than to the non-normative properties themselves, these properties, I believe, are irreducibly normative.

The situation with normativity is analogous to the situation with consciousness or phenomenal experience. Consciousness, I think we can say, is a form of organized matter, and phenomenal properties are equivalent to patterns in physical properties: properties which are not themselves phenomenal, but which, when they come together in a certain way, form phenomenal properties. Phenomenal properties (quite plausibly) cannot be completely reduced to physical properties, but they are nonetheless completely metaphysically dependent on them: they are substantive, but not fundamental, features of the universe. In the case of consciousness and phenomenal experience, to simply deflate the phenomenal into the physical, or to take the phenomenal to be somehow metaphysically primitive, or indeed to deny its existence altogether, is to avoid the crucial task of actually *understanding* consciousness by *explaining* exactly how the conscious arises from the non-conscious. Likewise, to deflate normativity into a function of predicates and concepts (or the like), as the reductionist does, or to regard it as a fundamental

feature of the universe, like the non-naturalist, or to simply deny its existence, like the error theorist, is to avoid the task of explaining normativity, and therefore of truly understanding it as I believe it deserves to be understood.

## 9. Conclusion

I have argued that there are mind-independent, irreducibly normative truths, and that these truths can be known on the basis of observation and inference to the best explanation. The Agony Principle, according to which we always have a reason to want and to try to avoid all future agony, is a prime example of an irreducibly normative truth that can be known empirically by inference to the best explanation. Normative truths do not always need to be known in this way, however. Parfit's Wager, according to which we have reasons to believe in (mind-independent) normative truths, and to engage in the form of deliberation which is committed to the existence of such truths (normative deliberation), is an example of a substantive normative truth which can be known *a priori*. But our ability to respond to normative reasons in a sensitive, counterfactually dependent manner suggests that normative facts and properties are genuinely causally efficacious (unless, that is, we interpret causality in an overly narrow sense, a sense that would actually rule out the possibility of facts and properties in general having any causal standing). And this in turn suggests that the normative domain is not autonomous from, but continuous with, the domain of the natural sciences, in the same way that other *a priori* accessible domains, such as logic and mathematics, are nonetheless perfectly continuous with science.

It is true that normative principles are not currently used in science in the way that logical and mathematical principles are. But since, returning to the definition of normativity with which we started, normativity is a force that applies specifically to the decisions and responses of conscious agents, and no one can justifiably claim that we fully understand these kinds of responses in a scientific way, I think there is still very much room for irreducibly normative principles to feature in a future “science of man,” which will not only explain why we respond to the world in the ways that we do, but which will also help us to understand which kinds of responses the world actually calls for. I believe the view I have advocated provides us with the most reliable basis for expanding our normative knowledge and for reaching greater convergence on our normative judgements.

### 9.1. *Robust Metanormative Naturalism and the Holy Grail of Metaethics*

I will end this dissertation by briefly considering the prospects for Robust Metanormative Naturalism to provide a basis not only for normative objectivity, but also for a stronger form of *ethical* objectivity, construed so as to cover specifically moral and altruistic reasons rather than just normative reasons in general. If the arguments I have made are correct, they show that we have normative reasons to be concerned with the contents of *our own* conscious experiences (such as agony) as individual conscious agents, but does this mean that we also have normative reasons to be concerned with the conscious experiences of others? Does the Agony Principle mean that Alan Gibbard’s ideally coherent Caligula, who “values the activity of

torturing others for fun, is perfectly logically and instrumentally consistent in holding this value, and is making no mistakes about the non-normative facts,” (Street 2016b, 164) nonetheless has sufficient normative reason not to torture others for fun (because of the intrinsic badness of the suffering this will cause, or for any other reason for that matter)? Many of us have a strong inchoate sense that such an agent is badly mistaken in thinking that the suffering of others has no bearing on how he ought to act. As Street explains,

The Holy Grail in secular metaethics would be a vindication of this inchoate sense—a philosophical account that clearly explained, with no mysterious metaphysical or epistemological remainder, and in a way that rang deeply true with our pre-philosophical sense of things, exactly *what* such an agent’s mistake would consist in when he or she contemplated the suffering of others and saw it as counting for nothing. (165)

Parfit’s Wager and the Normative Explanations argument, which together yield the Agony Principle, meaning that we always have a reason to (want to) avoid agony that comes not from our contingent desires, but from the intrinsic phenomenal character of agony itself, are, I think, both necessary steps on the road to discovering this Holy Grail. But the Agony Principle on its own does not tell us that we have reasons to be concerned with anyone else’s suffering. A further step is needed, therefore, to reach the conclusion that even an ideally coherent Caligula is mistaken when he contemplates the suffering of others and sees it as counting for nothing.

The step that Street herself tries to take to reach this Holy Grail is to observe that all agents face what she calls the “*problem of attachment and loss*” (161), which is simply the fact that “[e]ach of us faces the inevitable loss of everyone and everything

we love,” and that “[m]ore generally, we face an ineliminable gap between how things are and how we would like them to be” (163). Street suggests that, for every one of us, the optimal solution to this problem involves adopting what she calls the “*standpoint of pure awareness*” in which one identifies with “a universal point of view on the world that transcends any particular, finite point of view ... a point of view which is not itself vulnerable to loss” (186). She then goes on to suggest that “the standpoint of pure awareness is also an ethical standpoint,” because “[t]his point of view sees itself everywhere and is necessarily compassionate with the lived experience of every being. It recognizes that there is no real difference between ‘self’ and ‘other’—that instead ‘I’ am everywhere, in ‘you’ as much as ‘me’” (186-87). In other words, if we all have normative reasons to be concerned with our own conscious experiences, and consequently to (want to) avoid extremely negative experiences like agony and loss, we all have normative reasons to adopt the standpoint of pure awareness: adopting this standpoint is the only way we can avoid being vulnerable to loss. And if we all have reasons to adopt the standpoint of pure awareness, we all have reasons to have the same kind of concern with the conscious experiences of every other being as we do with our own conscious experiences (because this is what the standpoint of pure awareness necessarily involves). Even Caligula, therefore, actually has most reason to be compassionate towards others rather than torturing them for his own amusement.

Thus outlined, this particular map to the Holy Grail does not strike me as totally coherent. If one identifies with a standpoint that takes one from being

primarily concerned with one's own lived experience (or with the experience of a small minority of people) to being far more substantially concerned with the lived experience of all conscious beings, then surely one's vulnerability to loss has thereby only *increased*, rather than being somehow removed or reduced. If being concerned with my own experience makes me vulnerable to loss in the first place, becoming equally concerned with the experience of *all* conscious beings will surely make my vulnerability to loss equivalent, not to the absence of any such vulnerability, but rather to the combined vulnerability of all conscious beings. I will therefore be far less vulnerable to loss, *not* if I am equally compassionate towards the suffering of all conscious beings, but in fact if I am strictly concerned only with my own suffering and well-being. I have far less to lose as just one conscious being than as the sum of all conscious beings and their combined capacity for suffering. Parfit (2011b) interprets Nietzsche as claiming that "[t]o live cheerfully and with a good conscience ... each of us needs to have some *horizon* around ourselves ... If we were fully aware of the sufferings of others, we would be 'overwhelmed by compassion,' and not want to live" (574). For the same reason, I don't think it is viable to maintain that we have reasons to be compassionate towards all conscious beings that ultimately stem from our self-interest in solving a problem that applies to us as individuals.<sup>187</sup>

---

187. Moreover, because Street also argues that the problem of attachment and loss is a consequence of any evaluative point of view whatsoever, no matter the particulars (such as whether or not it values compassionate concern for all conscious beings) (2016b, 175), and it certainly seems that it must be such a consequence, since to have any evaluative point of view on the world is to encounter the fact that the world will never fully live up to one's values, it is difficult to see how the standpoint

But I think there is still something in the idea that there is no real difference between “self” and “other” (that instead “I” am everywhere, in “you” as much as in “me”), and that this, combined with the idea that there are normative reasons which ultimately stem, not from our desires, but from the intrinsic phenomenal characters of our experiences, is what can finally lead us to that elusive Holy Grail. Robust Metanormative Naturalism tells us that some conscious experiences are inherently reason-giving (they have *ought-to-be-avoidedness* or *ought-to-be-pursuedness* built into them). If we combine this metanormative view with the right view of personal identity and the self—a view on which personhood and the self are, on the one hand, “thick” enough for there to be reasons *for* persons to respond in certain ways in the first place, and yet at the same time also “thin” enough to block the relativization of the normative reasons that experiences like agony generate to the persons who stand to have the particular experiences in question—this should entail that we have normative reasons to be concerned not only with our own conscious experiences, but also with the conscious experiences of others. If this view of personhood is correct, Caligula would have reasons to cease his practice of torturing others for fun, reasons which come directly from the nature of the suffering he is causing, because the separateness of his own personhood and that of his victims would not be robust enough to prevent the reason-giving force of their suffering from transmitting to

---

of pure awareness could possibly be an ethical (and thus evaluative) standpoint at all if it is also supposed to be a point of view which is itself not vulnerable to loss.

him. His mistake would then consist precisely in thinking and acting as if the difference or boundary between himself and others is more real than it really is.

This idea is to some extent a familiar one. If one has a reason to (want to) avoid all future agony, one has a reason to prevent one's future self from being in agony. If one has a reason to be concerned with the conscious experiences of one's future self, one thereby has a reason to be concerned with the experiences of a self that is distinct from one's present self. Prudential reasons are therefore reasons to be concerned with the suffering and well-being of a self which is in an important sense "other" than the present self which actually has these reasons. And altruistic reasons are essentially reasons of this same kind: reasons to be co concerned with the suffering and well-being of selves other than the self which has actually these reasons. If there is no robust difference between one's future self and another self which exists outside one's own body, then if one has reasons to prevent one's future self from suffering, it arguably follows that one also has reasons to prevent the suffering of other selves which are distinct both from one's present self *and* one's future self. Christine Korsgaard (2009) argues that, in light of the choice-dependent nature of personal identity—in her words, "what counts as me, my incentives, my reasons, my identity, depends on, rather than proceeds, the kinds of choices that I make" (199)—interacting with others is rationally and normatively equivalent to interacting with yourself: "the requirements for unifying your agency internally are the same as the requirements for unifying your agency with that of others" (202). For Korsgaard, you cannot act for reasons at all unless you actively *constitute* yourself as

a person and unify your agency internally (204), which means that the reasons that apply to your future self must also have normative force for your present self, and if you meet the requirements for doing this, you also meet the requirements for unifying your agency with that of others, which means that the reasons that apply to them must also have normative force for you.

I would argue that the above argument fails within Korsgaard's own constructivist framework<sup>188</sup>: if there are no categorical reasons, I don't think even the reasons of our own present selves are normative for us, let alone the reasons of our future selves or the reasons of other selves that exist outside our own bodies. In this scenario, we may have hypothetical reasons to take the reasons of these other selves into account, but on my understanding at least, these would not be truly normative reasons. Yet if we are Robust Metanormative Naturalists, we can reject metaethical constructivism while still retaining its crucial insights about the nature of personhood and its significance for personal and interpersonal interaction. If Korsgaard is right that what counts as me depends on, rather than precedes, the kinds of choices that I make, then my personhood will be real enough for me to have normative reasons for action in the first place, but not real enough for it to be legitimate to relativize the Agony Principle to mean that I only have reasons to (want to) avoid *my own* future agony: the inherently reason-giving nature of agony will always precede the construction of my personhood that is based inescapably on the

---

188. I take it that metaethical constructivism is committed to the doctrine of Internalism (or subjectivism) about reasons, which I do not believe to be a coherent position.

choices that I make. It will thus provide a normative basis for me to constitute my identity in one way rather than other, and thus in a way that involves my having a concern for the suffering and well-being of others, rather than just my own.<sup>189</sup>

Parfit believed that our continued existence and identity as persons is not a deep further fact, distinct from our physical and psychological continuity, nor a fact that must be all-or-nothing, in the sense that each of us always either is or is not one particular person at any given time, but is instead simply certain direct connections between past, present and future experiences, such as those involved in experience-memory, or in the carrying out of an earlier intention. In a famous passage (1984), he describes what happened to him when he came to accept this “Reductionist View” of the self and personal identity:

When I believed that my existence was a such a further fact, I seemed imprisoned in myself. My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air. There is still a difference between my life and the lives of other people. But the difference is less. Other people are closer. I am less concerned about the rest of my own life, and more concerned about the lives of others. (281)

If there is no glass wall between oneself and others, and one’s conscious experiences give one normative reasons to be concerned about one’s own life, then these same

---

189. This may also serve to vindicate the oft-maligned move that Mill (1998) tried to make from the premise “that each person’s happiness is a good to that person” to the conclusion that “the general happiness, therefore, [is] a good to the aggregate of all persons” (81). If personal identity is ultimately choice-dependent, the reason we have to pursue happiness *as such* would already exist prior to the reasons we have to pursue *our own* happiness (which would come into existence only after we construct our separate individual identities). We could not then easily appeal to the “separateness of persons” in order to block Mill’s conclusion.

experiences should also give others normative reasons to be concerned about one's life, and the experiences of others should give one normative reasons to be concerned about their lives as well. If what we essentially are is just experiences that are related to each other in certain (direct) ways, and thus we are not really imprisoned inside glass tunnels that robustly separate us from each other, the normative reasons these experiences generate are likewise not imprisoned within the walls of glass tunnels, unable to transmit their normative force from one person to the other. The best explanation for why Parfit became more concerned with the lives of others upon accepting the Reductionist View could be that he became more sensitive to the normative reasons generated by the experiences of others (experiences that just happen not to be related in certain ways to certain other experiences), because he was longer imprisoned in the illusory glass tunnel of the self that kept him from responding appropriately to these normative reasons. And it may be the fact that agents like Caligula are so deeply imprisoned within this false tunnel—which in their case seems darker than glass—preventing them from responding to the reasons given to them by the sufferings of others, which ultimately explains their malignant behaviour. I will not defend this view of personal identity here; I will simply remark that, if something similar to this view turns out to be true, then given the resources provided by Robust Metanormative Naturalism as I have defended it, the Holy Grail of Metaethics may yet be firmly within reach.

## References

- Armstrong, D. M. *Truth and Truthmakers*. Cambridge: Cambridge University Press, 2004.
- Armstrong, D. M. *Sketch for a Systematic Metaphysics*. Oxford: Oxford University press, 2010.
- Blackburn, Simon. *Oxford Dictionary of Philosophy*, 3<sup>rd</sup> Edition. Oxford: Oxford University Press, 2016.
- BonJour, Laurence. *In Defense of Pure Reason: A Rationalist Account of A Priori Justification*. Cambridge: Cambridge University Press, 1998.
- Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press, 1996.
- Copp, David. *Morality, Normativity, and Society*. Oxford: Oxford University Press, 1995.
- Copp, David. *Morality in a Natural World: Selected Essays in Metaethics*. Cambridge: Cambridge University Press, 2007.
- Copp, David. "Normative Naturalism and Normative Nihilism." In *Reading Parfit: On What Matters*, 28-53. Edited by Simon Kirchin. Oxford: Routledge, 2017.
- Cuneo, Terrence and Shaffer-Landau, Russ. "The moral fixed points: new directions for moral nonnaturalism." *Philosophical Studies* 171, no. 3 (December 2014): 399-343. doi: 10.1007/s11098-013-0277-5.
- Dennett, Daniel C. *Consciousness Explained*. London: Penguin, 1991.
- Dodd, Julian. "Farewell to States of Affairs." *Australasian Journal of Philosophy* 77, no. 2 (1999): 146-160. doi: 10.1080/00048409912348901.
- Draper, Paul. "God, Science and Naturalism." In *The Oxford Handbook of Philosophy of Religion*, 272-303. Edited by William J. Wainwright. Oxford: Oxford University Press, 2007.

- Dworkin, Ronald. "Objectivity and Truth: You'd Better Believe It." *Philosophy & Public Affairs* 25, no. 2 (Spring 1996): 87-139.  
<https://www.jstor.org/stable/2961920>.
- Enoch, David. "An Outline of an Argument for Robust Metanormative Realism." In *Oxford Studies in Metaethics: Volume 2*, 21-50. Edited by Russ Shafer-Landau. Oxford: Oxford University Press, 2007.
- Enoch, David. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press, 2013.
- Evers, Daan. Review of *Moral Error Theory: History, Critique, Defence*, by Jonas Olson. September 2014. <https://ndpr.nd.edu/news/moral-error-theory-history-critique-defence/>.
- Fitzpatrick, William J. "Ontology for an Uncompromising Ethical Realism." *Topoi* 37, no. 4 (November 2016): 537-547. doi: 10.1007/s11245-016-9443-7.
- Hájek, Alan. "Pascal's Wager." In *The Stanford Encyclopedia of Philosophy*. Summer. Edited by Edward N. Zalta. 2018.  
<https://plato.stanford.edu/archives/sum2018/entries/pascal-wager/>.
- Harman, Gilbert. "Quine on Meaning and Existence, I. The Death of Meaning." *The Review of Metaphysics* 21, no. 1 (September 1967): 124-151.  
<https://www.jstor.org/stable/20124498>.
- Harman, Gilbert. "Is There a Single True Morality?" In *Explaining Value and Other Essays in Moral Philosophy*, 77-100. Oxford: Oxford University Press, 2000.
- Jackson, Frank. "In Defence of Reductionism in Ethics." In *Does Anything Really Matter? Essays on Parfit on Objectivity*, 195-211. Edited by Peter Singer. Oxford: Oxford University Press, 2016.
- Joyce, Richard. *The Myth of Morality*. Cambridge: Cambridge University Press, 2001.
- Kahane, Guy. "Feeling Pain for the Very First Time: The Normative Knowledge Argument." *Philosophy and Phenomenological Research* 80, no. 1 (2010): 20-49.  
<https://www.jstor.org/stable/40380500>.
- Kahane, Guy. "If Nothing Matters." *Noûs* 51, no. 2 (2017): 327-353. doi: 10.1111/nous.12146.

- Korsgaard, Christine M. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009.
- Leiter, Brian. "Moral Facts and Best Explanations." In *Naturalizing Jurisprudence: Essays on American Legal Realism and Naturalism in Legal Philosophy*, 203-224. Oxford: Oxford University Press, 2007. doi: 10.1093/acprof:oso/9780199206490.001.0001.
- Leiter, Brian. "Normativity for Naturalists." *Philosophical Issues* 25, Normativity (2015): 64-79. doi: 10.1111/phis.12046.
- Lutz, Matt and Finlay, Stephen. Review of *Moral Error Theory: History, Critique, Defence*, by Jonas Olson. *Ethics* 125, no. 4 (July 2015): 1219-1225. <https://www.jstor.org/stable/10.1086/680900>.
- Mackie, J. L. *Ethics: Inventing Right and Wrong*. London: Penguin Books, 1990.
- McPherson, Tristram. "Against Quietist Normative Realism." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 154, no. 2 (June 2011): 223-240. <https://www.jstor.org/stable/41487663>.
- Mill, J. S. *Utilitarianism*. Edited by Roger Crisp. Oxford: Oxford University Press, 1998.
- Mills, Eugene. "Interactionism and Overdetermination." *American Philosophical Quarterly* 33, no. 1 (January 1996): 105-117. <https://www.jstor.org/stable/20009850>.
- Olson, Jonas. "In Defense of Moral Error Theory." In *New Waves in Metaethics*, 62-84. Edited by Michael Brady. London: Palgrave Macmillan, 2011.
- Olson, Jonas. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press, 2014.
- Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.
- Parfit, Derek. *On What Matters: Volume One*. Oxford: Oxford University Press, 2011a.
- Parfit, Derek. *On What Matters: Volume Two*. Oxford: Oxford University Press, 2011b.
- Parfit, Derek. *On What Matters: Volume Three*. Oxford: Oxford University Press, 2017.

- Quine, W. V. O. "Necessary Truth." In *Ways of Paradox*, 68-76. Cambridge, Massachusetts: Harvard University Press, 1976.
- Quine, W. V. O. "On the Nature of Moral Values." *Critical Inquiry* 5, no. 3 (Spring 1979): 471-480. <http://www.jstor.org/stable/1342996>.
- Quine, W. V. O. "Two Dogmas of Empiricism." In *From a Logical Point of View: Nine Logico-Philosophical Essays*, 20-46. Cambridge, Massachusetts: Harvard University Press, 1980.
- Robb, David and Heil, John. "Mental Causation." *The Stanford Encyclopedia of Philosophy*. Summer. Edited by Edward N. Zalta. 2019. <https://plato.stanford.edu/archives/sum2019/entries/mental-causation/>.
- Robinson, Howard. "Dualism." *The Stanford Encyclopedia of Philosophy*. Fall. Edited by Edward N. Zalta. 2017. <https://plato.stanford.edu/archives/fall2017/entries/dualism/>.
- Russell, Bertrand. *History of Western Philosophy*. London: Routledge, 2004.
- Scanlon, T. M. "Rawls on Justification." In *The Cambridge Companion to Rawls*, 139-168. Edited by Samuel Freeman. Cambridge: Cambridge University Press, 2003.
- Scanlon, T. M. *Being Realistic About Reasons*. Oxford: Oxford University Press, 2014.
- Sher, Gila. "The Foundational Problem of Logic." *The Bulletin of Symbolic Logic* 19, no. 2 (June 2013): 145-198. <http://www.jstor.org/stable/41955403>.
- Skorupski, John. *The Domain of Reasons*. Oxford: Oxford University Press, 2010.
- Sosa, Ernest. "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge." *Midwest Studies in Philosophy* 5, no. 1 (1980): 3-26. doi: 10.1111/j.1475-4975.1980.tb00394.x.
- Street, Sharon. "Constructivism in Ethics and the Problem of Attachment and Loss." *Aristotelian Society Supplementary Volume* 90, no. 1 (June 2016a): 161-189. doi: 10.1093/arisup/akw002.
- Street, Sharon. "Nothing 'Really' Matters, but That's Not What Matters." In *Does Anything Really Matter? Essays on Parfit on Objectivity*, 121-148. Edited by Peter Singer. Oxford: Oxford University Press, 2016b.

- Street, Sharon. "Objectivity and Truth: You'd Better Rethink It." In *Oxford Studies in Metaethics: Volume 11*, 293-333. Edited by Russ Shafer-Landau. Oxford: Oxford University Press, 2016c. doi: 10.1093/acprof:oso/9780198784647.003.0012.
- Streumer, Bart. *Unbelievable Errors: An Error Theory about All Normative Judgements*. Oxford: Oxford University Press, 2017.
- Sturgeon, Nicholas L. "Moral Explanations." In *Essays on Moral Realism*, 229-255. Edited by Geoffrey Sayre-McCord. London: Cornell University Press, 1988.
- Temkin, Larry S. "Has Parfit's Life Been Wasted? Some Reflections on Part Six of *On What Matters*." In *Does Anything Really Matter? Essays on Parfit on Objectivity*, 1-34. Edited by Peter Singer. Oxford: Oxford University Press, 2016.
- Unger, Roberto Mangabeira and Smolin, Lee. *The Singular Universe and the Reality of Time*. Cambridge: Cambridge University Press, 2015.
- Wedgwood, Ralph. *The Nature of Normativity*. Oxford: Oxford University Press, 2007.
- Young, James O. "The Coherence Theory of Truth." *The Stanford Encyclopedia of Philosophy*. Fall. Edited by Edward N. Zalta. 2018.  
<https://plato.stanford.edu/entries/truth-coherence/>.
- Yudkowsky, Eliezer. "What is Evidence?" *Less Wrong* (blog). September 22, 2007a.  
<https://www.lesswrong.com/posts/6s3xABaXKPdFwA3FS/what-is-evidence>.
- Yudkowsky, Eliezer. "A Priori." *Less Wrong* (blog). October 8, 2007b.  
<https://www.lesswrong.com/posts/qmqLxvtsPzZ2s6mpY/a-priori>.