HOW TO BE (AND HOW NOT TO BE) A NORMATIVE REALIST

David Faraci

A Dissertation
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2012

Committee:
David Shoemaker, Advisor
Anne K. Gordon
Graduate Faculty Representative
Christian Coons
David Copp
Tristram McPherson
Sara Worley
ABSTRACT

Broadly and somewhat roughly speaking, metanormative theorists who maintain that there is normative truth fall into one of three camps: non-naturalist, naturalist and expressivist. I am interested in the prospects for normative truth, and thus in which, if any, of these positions offers hope for the discovery of such truth. In each of three chapters, I address one of these views. I conclude that our best hope is a view most closely related to naturalism (though I reject this classification for one that I believe better captures what is at stake in the literature I focus on).

In Chapter 1, I target expressivism, according to which normative thought and language express non-cognitive attitudes. I explain why it will be difficult, if not impossible, for expressivists to account for a kind of commonplace nihilistic doubt that, I argue, is a symptom of the widely accepted “objective” nature of normativity. I conclude that an inability to account for such doubt should be considered a serious problem for expressivism.

In Chapter 2, I address a prominent form of non-naturalism, according to which the normative is “metaphysically autonomous”—neither identical with, nor constituted by, nor constitutive of anything non-normative. I examine several important explanatory challenges non-naturalists face in normative epistemology, metaphysics and semantics. I argue that it will be difficult, if not impossible, to develop their view such that it meets these challenges while remaining plausible.

The Open Question Argument and its contemporary cousin, Normative Twin Earth, are the most prominent objections to naturalism. In Chapter 3, I argue that because these arguments rely on semantic rather than metaphysical intuitions, they should be understood as targeting semantic rather
than metaphysical views (e.g., not naturalism). I propose that the appropriate targets are views committed to a particular class of *reference-fixing relations* for normative terms. I then examine the prospects for developing a view along these lines that does not fall prey to the objections mentioned. I conclude that there is hope for realism of this kind, and close with a rough sketch of how such a view might be developed more fully.
ACKNOWLEDGEMENTS

I am extraordinarily grateful to the members of my committee, Christian Coons, David Copp, Tristram McPherson, David Shoemaker and Sara Worley, for their support and guidance throughout this process. Among these, I am particularly grateful to Christian Coons and Tristram McPherson. Christian has been a constant source of inspiration over the five years I have known him, and has worked more than anyone could reasonably expect to help me carve out a place for myself in this discipline. I have known Tristram for a comparatively short period of time; but during that time he has helped, supported and influenced me and my work to an extent that would be impressive even if he weren’t at a different institution entirely.

Additional thanks go to: R.G. Frey (who, regretfully, had to excuse himself from my committee for health-related reasons) for his help and guidance, especially early in my graduate career. Several people who participated in a useful discussion of the arguments in Chapter One on the blog PEA Soup (and, in some cases, via email), especially Jamie Dreier, Alex Gregory, Kate Manne, Mark Schroeder and Jussi Suikkanen. David Enoch (whose work, the quality of which is something to aspire to, is a focal point of my own) for reading and offering helpful and insightful comments on Chapter Two. Michael Weber, who, as chair of the department, has been unwavering in his support. Margy DeLuca and Priscilla Ibarra, our department secretaries and the true power behind the throne. Jacob Sparks and Mark Wells, two of my colleagues in the graduate program at Bowling Green with whom I have had numerous useful discussions about my research. Ruth Chang and Larry Temkin, my unofficial undergraduate advisors at Rutgers University, without whose encouragement I would not be here today. And my parents, Robert Faraci and Claire Scarano, for always encouraging me to pursue my passion for this work, despite the danger that it would be utterly useless as a source of income.
# TABLE OF CONTENTS

INTRODUCTION .................................................................................................................................................. 1

CHAPTER ONE. ENEMY OF MY ENEMY: NIHILISM AND THE CASE FOR COGNITIVISM ........................................... 6

CHAPTER TWO. CAN MOOREAN REALISM SHOULDER ITS EXPLANATORY BURDEN? ........................................ 32

CHAPTER THREE. HOW TO BE A NORMATIVE REALIST ..................................................................................... 64

BIBLIOGRAPHY .................................................................................................................................................. 98
INTRODUCTION

Broadly and somewhat roughly speaking, metanormative theorists who maintain that there is normative truth fall into one of three camps. First, there are those who believe that the normative is metaphysically autonomous, that normative facts and properties are unlike those of any other kind. Typically, these theorists are called non-naturalists, though some go by other titles (which may or may not track substantive disagreements within the camp), such as “robust realists” or “quietists.”

Second, there are those who believe that normative truth is just another part of the natural world, that what’s good might turn out to just be what’s happiness-maximizing, or what we would all consent to, or whatever. Typically, these are called naturalists, though they also go by other titles (which again may or may not track substantive disagreements within the camp), such as “reductive realists” or “constructivists.”¹

Finally, there are those who believe in normative “truth” only in an extended sense. They believe that normative language expresses non-cognitive, “desire-like” (as opposed to cognitive, “belief-like”) attitudes, but that nevertheless they can “earn the right” to talk about normative claims’ being true or false—typically through appeal to a deflationary account of truth. These are typically called non-cognitivists, expressivists, or quasi-realists (going from least to most specific in terms of their commitments).

¹ Some constructivists might object to being lumped in with the naturalists. However, there are some good arguments, I think, that constructivism is not a genuine metanormative alternative at all, and that to the extent that it is, would be nearly indistinguishable from naturalism. See, e.g., Enoch (2009). Whether or not this is correct, I am confident that the arguments to come rely on nothing of substantive difference between these views.
I am interested in the prospects for normative truth, and thus in which, if any, of these positions offers hope for the discovery of such truth. In each of three chapters, I address one of these views. I conclude that our best hope is a view most closely related to naturalism (though I reject this classification for one that I believe better captures what is at stake in the literature I focus on). In the remainder of this introduction, I say a bit more about each of the arguments to come, and highlight some areas where further research is warranted.

1. Expressivism

Expressivists hold that normative language expresses non-cognitive, “desire-like” attitudes or mental states, as opposed to cognitive, “belief-like” ones. Many people hold that normative thought and language concern normative truth—something that is external to us in much the way that truths about chairs and tables is, or perhaps better, truths about mathematics. Even the philosophically ignorant can see that there is tension between these two views. If I tell you that all thought and language about something is just an expression of a desire-like attitude, it is hard to see how that thing could be appropriately “objective.”

Perhaps, though, I shouldn’t say “even” the philosophically ignorant, because a number of the philosophically sophisticated—most notably quasi-realists like Simon Blackburn (e.g., 1998) and Alan Gibbard (e.g., 2003)—claim that their expressivism is compatible with a kind of normative “truth.” Whether or not this is right, in the first chapter I argue that there are reasons to think that expressivism is incompatible with the objectivity of such truth—with the claim that normative thought and language permit us to think and talk about whether anything matters though we might well know that many things matter to us. Specifically, I argue that expressivism likely cannot account for a kind of commonplace nihilistic doubt that is a symptom of this feature of normative thought and language and that this should be considered a serious problem for the view.
2. Non-Naturalism

G.E. Moore (1903) famously offered the Open Question Argument (OQA) for non-naturalism. Consider the question: “I know that X is pleasurable, but is it good?” Regardless of whether you think that being pleasurable is a good-making feature of things, this question has an “open” feel, in that X’s being pleasurable does not, all by itself, seem to settle X’s being good (in the way that being a bachelor seems to settle, all by itself, being unmarried). This is evidence that the connection between pleasure and goodness is not a conceptual one, and thus that an analysis of goodness cannot be given in terms of pleasure. What’s more, Moore thought that the same could be said for any pair of normative/non-normative concepts. Ultimately, Moore concluded from this that normative properties are not identical to any natural properties. He also concluded that normative properties are autonomous or sui generis—that they are unlike and metaphysically independent of properties of any other kind.

Typically, those who share Moore’s conclusions have been called non-naturalists. For a variety of reasons I discuss in the text, this label is problematic. Given this, I take as my target “Moorean realism,” which I stipulate to be the view that the normative is metaphysically autonomous combined with two theses that nearly all non-naturalists accept: (a) that our normative attitudes cannot ultimately explain the normative facts (mind/attitude/response/stance-independence) and (b) that the normative is causally inefficacious. (It is worth noting that, arguably, not only do those who maintain normative autonomy typically accept the causal inefficacy and mind-independence of the normative, but that they are in fact committed to such acceptance. While I do not argue for such commitments here, this is a possible avenue for future research.)

Mooreans face a number of important explanatory challenges in normative metaphysics, epistemology and semantics. I set out to assess their ability to meet these challenges and conclude
that it will be difficult, if not impossible, for them to develop their view such that it can both meet these challenges and remain plausible.

One further point: I think that my arguments in Chapter 2 make it clear that, dialectically speaking, Mooreans are in a tough spot. Even they, I think, would have to agree that the stories they tell to save their view are a bit of a stretch. They maintain their view, though, often because (apparently) it seems to them that Moorean realism is the only genuine alternative to error theory—the view that though normative thought and language purport to be about normative facts, there are no such facts. Indeed, I suggest that given the nature of the Moorean view, it could be defensible only if it were the only realist alternative. In Chapter 3, I offer what I think may well be a viable alternative, and explain how, if successful, it would retain the advantages of Moorean realism without many of the costs. If that view is indeed successful, the Moorean view is arguably untenable.

3. Naturalism

To say that the normative is non-natural is to make a metaphysical claim. But the only argument for the view discussed thus far—the OQA—is a semantic argument. While relevant inferences from semantics to metaphysics made sense in Moore’s time, developments in both the philosophy of language and metaphysics have rendered such inferences highly controversial. Interestingly, the most prominent contemporary objection to naturalism—Horgan and Timmons’ (e.g., 1991 and 1992) Normative Twin Earth (NTE)—likewise relies on semantic intuitions; indeed, it can plausibly be thought of as a descendent of the OQA. Arguably, NTE licenses metaphysical conclusions in a way that the OQA does not. But only arguably; the assumptions one needs to make to get from semantics to metaphysics remain controversial. In Chapter 3, I propose that we can avoid controversial assumptions by carving up the metanormative landscape in terms of metasemantics rather than metaphysics.
Having drawn the appropriate distinctions, I consider the force of the OQA and the newer NTE against what I call non-Moorean realism—a particular view about the nature of reference-fixing relations for normative terms. I conclude that the OQA and NTE raise serious challenges to development of a viable non-Moorean view, but that hope remains. The close of the dissertation is a rough and exploratory sketch of how to develop a substantive non-Moorean view. If I am right, pursuing the non-Moorean line is our best hope for developing a metanormative theory on which there is genuine, objective, normative truth. This last is, of course, one of the largest areas for further research in the dissertation. In future, I hope to more fully explore non-Moorean realism’s potential.

---

2 In Chapter 2, I discuss a metaphysical view I call Moorean realism. In Chapter 3, I discuss a metasemantic view I call non-Moorean realism. As the chapters of the dissertation are fairly self-contained, I do not address this shift from metaphysics to metasemantics in the text (with the exception of a footnote in Chapter 3). In any case, given the state of the literature, the conflation is of little importance. With the exception of David Enoch (2011), to the best of my knowledge no (metaphysical) Moorean has said much about how they mean to characterize the reference-fixing relation for normative terms. And Enoch’s view is apparently consistent with both the metasemantic and metaphysical understandings of Moorean realism. Indeed, so far as I am aware, the only person whose view’s classification might be affected by this shift is Ralph Wedgwood (2007) (from whom Enoch draws much of his account). Arguably, Wedgwood should be considered a metasemantic Moorean but a metaphysical non-Moorean. Obviously, getting clear on what exactly the possible combinations of semantic and metaphysical view are in this area is another avenue for potential future research.
CHAPTER ONE

ENEMY OF MY ENEMY: NIHILISM AND THE CASE FOR COGNITIVISM

To many, expressivism—according to which normative thought and language express non-cognitive attitudes—seems like a non-starter. If all goes well when we deliberate, our actions are informed by our normative attitudes, which are themselves informed by the normative truth. Yet according to expressivism, ultimately there is nothing to normativity beyond such attitudes. The expressivist “has to believe that [normativity] somehow depends on us, that the ultimate explanation of why it is that certain [normative] claims are true has something to do with us and our feelings and attitudes” (Enoch 2011, 36). And so a key piece of the story seems missing—our attitudes may inform our actions, but there is nothing “out there” to inform those attitudes, even if, as some maintain, there could still be normative “truth.”¹ In an important sense, whether or not expressivism is anti-realist, it seems anti-objectivist.

Most of the literature attacking expressivism focuses not on worries about objectivity, but on instances of the so-called embedding problem (EP).² Here, cognitivists charge that expressivism cannot account for particular aspects of the logical structure of normative thought and language; expressivists respond with an account; cognitivists either challenge it or move on to a new structural feature. No one budges. The situation is reminiscent of the debate over intelligent design, with its champions challenging evolutionists to explain an ever-evolving (no pun intended) parade of natural phenomena that seem “irreducibly complex” in order to prove that natural selection alone can’t explain the apparent order in the natural world.

---

¹ Notably quasi-realists like Simon Blackburn (e.g., 1998) and Alan Gibbard (e.g., 2003).

² One aspect of which is the more commonly referenced “Frege-Geach problem.”
Given this, it is tempting to set aside EP in favor of entirely new, objectivity-based lines of argument against expressivism, much as David Enoch does after concluding: “If God whispers in the ears of all cognitivists that [EP] can be very neatly solved, I do not see a trend of conversion to expressivism” (2011, 9–10). But even if we agree that EP alone is not enough to move the dialectic forward, we should not abandon it too quickly. While worries about embedding are not themselves worries about objectivity, I contend that EP can be employed to make the intuitive, objectivity-based case against expressivism more powerful. Specifically, I argue that expressivists are likely unable to both solve EP and make intelligible a kind of commonplace nihilistic doubt that is a symptom of the objectivist nature of normative thought and language.

1. The Negation Problem

In the discussion to follow, I draw heavily on Mark Schroeder’s (2008b) work on the “negation problem”—a particularly worrying aspect of EP—both in terms of his presentation of the problem and his proposed solution. I should note here that Schroeder does not merely propose a solution to the negation problem, but argues that his is the only plausible solution available to the expressivist. If that’s correct—as I believe it is—the arguments herein address expressivism generally. If Schroeder
(and I) are wrong about this, then some semantic views available to expressivists may not run up against the objectivity-based objections I raise.\footnote{However, I do not believe that my objections are just objections to Schroeder’s proposed expressivist semantics. I am confident that they apply to a number of other proposals in the literature as well, though I do not have the space to address this issue here. I focus on Schroeder because, again, I agree with him that alternate expressivist views fall to the negation problem in the ways he lays out.}

Consider the sentences:

S1. Murdering is wrong.

S2. Murdering is not wrong.

S1 and S2 are inconsistent. It is natural to think that the reason why they are inconsistent is that their contents are inconsistent. It is also natural to think that if I were to assert both S1 and S2, I would be inconsistent, in the sense that I would be rationally criticizable. Furthermore, it seems clear that if I assert S1 and you assert S2, we disagree.\footnote{All assertions referred to should be presumed sincere.}

The cognitivist can easily explain why each of these thoughts is natural, perhaps even correct. First, it seems plausible that in general sentences are inconsistent when and because they have inconsistent contents, as S1 and S2 apparently do. Second, assertions of S1 and S2 express beliefs, respectively, in S1 and S2. And one is inconsistent (again, in the sense of being objectionably irrational) if one believes inconsistent things. Third, and similarly, when I assert S1 and you assert S2, we express our respective beliefs in S1 and S2. And if I believe something that is inconsistent with what you believe, we disagree.

The expressivist doesn’t have it so easy. First, she cannot accept that the reason S1 and S2 are inconsistent is that they have inconsistent propositional contents (‘murdering is wrong’ and ‘murdering is not wrong’, respectively), because for her, S1 doesn’t have propositional content in the
way that it does for the cognitivist. And so, Schroeder tells us, for the expressivist, S1 and S2 must be inconsistent because the *attitudes they express* are inconsistent. Of course, a cognitivist could accept this view, too; he could say that S1 and S2 are inconsistent *because* belief in S1 is inconsistent with belief in S2. But such a move seems unmotivated. Intuitively, at least, it gets things the wrong way around: It is the fact that P and ~P are in conflict that explains why belief in them is inconsistent, not the reverse. But such is the view left the expressivist: S1 and S2 are inconsistent because each expresses some non-cognitive attitude, and the attitude expressed by asserting S1 is inconsistent with that expressed by asserting S2 (39-41).

The expressivist also faces a challenge regarding the inconsistency of someone who *thinks* both S1 and S2. Cognitivists can maintain that such a person is inconsistent just because his beliefs are. This is because beliefs are what Schroeder calls *inconsistency-transmitting attitudes*: “An attitude A is *inconsistency-transmitting* just in case two instances of A are inconsistent just in case their contents are inconsistent” (43). But not all attitudes are inconsistency-transmitting. For example, wondering whether S1 is not inconsistent with wondering whether S2. (Likewise, if I wonder whether S1 and you wonder whether S2, we do not therefore disagree.) So the expressivist must show that the non-cognitive attitude(s) she appeals to in developing her view is more like a belief than a wondering in this way. Minimally, she must show that *some* non-cognitive attitudes are inconsistency-transmitting.7

Even assuming she can do this, things get worse for the expressivist. Return to the wrongness of murdering. As Schroeder (44-45) points out, there are a number of relevant conditions

---

7 Schroeder (41-43) suggests that this last challenge may be met without too much difficulty, especially given that, arguably, *intention* is an inconsistency-transmitting non-cognitive attitude.
you might be in with respect to the proposition that murdering is wrong, differencing only⁸ in where, if anywhere, we place a “not.” It might be that:

S3. You think that murdering is wrong.
S4. You do not think that murdering is wrong.
S5. You think that murdering is not wrong.
S6. You think that not murdering is wrong.

Cognitivists have no trouble making sense of these different conditions. They just take “think” to mean “believe”:

3C. You believe that murdering is wrong.
4C. You do not believe that murdering is wrong.
5C. You believe that murdering is not wrong.
6C. You believe that not murdering is wrong.

Following Schroeder, let’s just suppose for the moment that the expressivist holds that to think something is wrong is to DISAPPROVE of it, where DISAPPROVAL is a particular non-cognitive attitude.⁹ Now the expressivist tries to translate each of the states you might be in:

3E. You DISAPPROVE of murdering.
4E. You do not DISAPPROVE of murdering.
5E. You…???
6E. You DISAPPROVE of not murdering.

---

⁸ Ignoring the grammatically necessary “do” in S4.

⁹ Throughout this chapter, I use small caps to denote “special” non-cognitive attitudes. These attitudes are not necessarily those that we associate with the names given them. Thus, DISAPPROVAL may not be the attitude commonly referred to as “disapproval.” Rather, it is a stand-in for some expressivist-friendly wrongness-connoting non-cognitive attitude.
Apparently, the expressivist has no way of differentiating between merely not thinking that murdering is wrong (S4) and thinking that murdering is not wrong (S5). The reason for this is simple: For the cognitivist, in each case one bears or does not bear a particular cognitive attitude (belief) towards a predicative proposition, which has both a subject (murdering) and a predicate (is wrong). But for the expressivist, the non-cognitive attitude (DISAPPROVAL) is borne simply towards an action—murdering. And so the expressivist has one less “slot for the not.”

At this point, Schroeder (45-48) argues, the expressivist has two options. First, she can introduce a new attitude. She can say that the translation includes:

5E. You TOLERATE murdering.

The problem is that once the expressivist goes this route, she loses her ability to explain why thinking that murdering is wrong is logically inconsistent with thinking that murdering is not wrong. The former is an expression of DISAPPROVAL, the latter of TOLERANCE. Of course, it is natural to think that DISAPPROVAL and TOLERANCE are inconsistent, perhaps because they are inter-definable. But the expressivist cannot make use of this suggestion. Suppose she claims that to TOLERATE X is just not to DISAPPROVE of X. She is now right back where she started, because now ‘You TOLERATE murdering’ is the same as ‘You do not DISAPPROVE of murdering’. Since the latter is how she understood S4, she is no better off with respect to understanding S5. Without another slot for the not, the expressivist simply cannot fully capture the differences between S3, S4 and S5. So, ultimately, if the expressivist is to go this route, she must maintain that DISAPPROVAL and TOLERANCE are primitive non-cognitive attitudes that are logically inconsistent. But it seems she can offer no explanation for why they are inconsistent. And this is problematic. Schroeder offers the following analogy:

Suppose, for example, that someone tells you that when she uses the word ‘not’ or the prefix ‘im-’ immediately before ‘permissible’, they are not to be understood as
meaning what ‘not’ normally does. Instead, she says, she believes in distinct, unanalyzable, and non-interdefinable properties of permisibility and impermissibility. And then suppose that she tells you that she also believes that it is impossible—logically impossible—for something to be both permissible and impermissible. Finally, she tells you that what makes ‘murder is permissible’ and ‘murder is not permissible’ logically inconsistent is that ‘murder is not permissible’ just means ‘murder is impermissible’, which is logically inconsistent with ‘murder is permissible’. Similarly, she insists, what makes ‘murder is impermissible’ and ‘murder is not impermissible’ logically inconsistent is that ‘murder is not impermissible’ just means ‘murder is permissible’, which is logically inconsistent with ‘murder is impermissible’.

Surely she leaves something to be explained! Obviously, her view will be a bad view about permisibility and impermissibility unless it turns out that ‘permissible’ and ‘impermissible’ are inconsistent. But that does not mean that she is entitled to assume it, in order to explain why ‘murder is permissible’ and ‘murder is not permissible’ are inconsistent! On the contrary, her view seems to have written out of existence everything that could be used to explain why permisibility and impermissibility are incompatible. (47-48)

The only alternative, Schroeder (58-61) maintains, is to make another slot for the not. And the way to do this is to follow the cognitivist’s example by proposing the existence of a general non-cognitive attitude that can be borne towards something with two “slots,” rather than simply towards an action. Schroeder calls this umbrella attitude BEING FOR, and it allows the expressivist to understand S3-S6 as follows:

3F. You are FOR blaming for murdering.
4F. You are not FOR blaming for murdering.

5F. You are FOR not blaming for murdering.

6F. You are FOR blaming for not murdering.

This, Schroeder (61) says, is the solution to expressivism’s negation problem.10 There is simply no other way to explain why “murdering is wrong” is inconsistent with “murdering is not wrong” because, first, expressivists have to understand the inconsistency of these sentences in terms of the inconsistency of the attitudes expressed by them and, second, the inconsistency of those attitudes is only explicable if the attitude expressed by each proposition is an instance of the same general attitude-type, which we are calling BEING FOR.

Importantly, if Schroeder is right, expressivists will need to extend his BEING FOR solution to a view about the semantics of all declarative sentences. After all, it is perfectly acceptable to make claims that involve both normative and non-normative predicates, such as “The sky is blue and murdering is wrong.” And, as Schroeder (2008b, chap. 7) argues, the only way to understand such sentences in a way that will maintain their logical form is to understand belief in terms of BEING FOR (or the other way around, but of course that’s not an option for expressivists).

So, according to what Schroeder calls “biforcated attitude semantics” (BAS), the semantic value of any predicate is a relation11 (e.g., ‘is wrong’ means ‘blaming for’), and judging that

10 Formally speaking, that is. Of course, the particular details concerning the nature of the attitude and what it is borne towards is entirely up for debate. Also, note that all of this holds only for “pure” expressivist views that are wholly non-cognitivist. Certain “hybrid” or “tempered” forms of expressivism are not bound in these ways, as they allow cognitive normative attitudes to play a role in their theory. I do not discuss such views further here. For discussion, see, e.g., Horgan and Timmons (2006), Schroeder (forthcoming), and Toppinen (forthcoming).

11 Technically, it’s a pair of relations. But this is irrelevant for our purposes.
something is so predicated means committing oneself to bearing the relevant relation towards it (e.g., judging something wrong means committing oneself to blaming for it). Thus, in brief, “we interpret the property associated with each [declarative] sentence, ‘P’, as telling us what someone who [judges] that P is thereby committed to doing” (Schroeder unpublished manuscript, 1).

2. Enter the Nihilist

John is a normative nihilist. He does not think that anything is right or wrong, good or bad, virtuous or vicious; he does not think that there is anything one ought to do, or that one has normative reason to do. In short, John makes no “substantive” normative judgements (SNJ)—no judgements that any normative property is ever instantiated. Speaking with Jane, a BAS expressivist, John asserts: “Murdering is not wrong.” Jane concludes that John has expressed his BEING FOR not blaming for murdering.

On its face, this is not a very plausible interpretation of what John said. It just doesn’t seem true that in judging that murdering is not wrong John commits to not blaming for murdering. And this may have nothing to do with his being a nihilist. Arguably, even non-nihilists can take themselves to have reasons to blame for things that are not wrong, such as when they take themselves to have non-moral reasons for wanting to dissuade people from doing those things.

All this means is that ‘blaming for’ probably isn’t the semantic value of ‘is wrong’. But that shouldn’t be very surprising, nor is it really a problem for BAS, because this particular analysis is not at all central to Schroeder’s defense; ‘blaming for’ is just a placeholder for whatever view on the

---

12 Of course, some expressivists will deny that any judgements are about normative properties per se. In that case, we can just think of SNJs as judgements that entail the proper application of a normative predicate, or something like that.
nature of wrongness expressivists want to adopt. The same is true of BEING FOR: It is a placeholder for a very general non-cognitive attitude of judgement or commitment of some kind.

Nevertheless, there are some important things that we know about how BAS works that tell us what kinds of attitudes could and could not fill in these placeholders. ‘Murdering is wrong’ is clearly a SNJ.\(^{13}\) Thus, one thing we know is that if the semantic value of ‘is wrong’ is relation R, then ‘BEING FOR bearing relation R to X’ is a SNJ. (For the sake of simplicity, I will refer to ‘bearing relation R to’ as ‘Φ-ing’.\(^{14}\)) Another thing we know, assuming that we agree that John can coherently maintain that murdering is not wrong (more on this later), is that ‘BEING FOR not Φ-ing X’ is not a SNJ.

In the following section, I argue that it will be difficult, if not impossible, for expressivists to explain why one of these judgements is a SNJ while the other is not. Given this, expressivists are apparently left with only two options for interpreting the nihilist: Change their interpretation, or claim that the nihilist is FOR not blaming for murdering, but is therefore inconsistent qua nihilist. I address and reject these responses in §4-5. If my arguments are successful, then Jane has no way of

---

\(^{13}\) Again, if you are displeased with talk of properties, you can simply take this to mean that ‘is wrong’ is a normative predicate and thus that judging ‘murdering is wrong’ means judging that a normative predicate is properly applied to something. Also, it’s worth noting that, on some views, wrongness is part of the definition of murder. In that case, ‘murdering is wrong’ might not be a SNJ, any more than ‘all bachelors are unmarried’ entails that something is a bachelor. So I’m assuming here that it isn’t the case that murdering is wrong by definition. It should be clear that nothing rests on my talking about murder rather than any number of other actions in my examples.

\(^{14}\) In order to avoid confusion with the objection above, which was specifically about understanding wrongness in terms of blaming, I will continue throughout this chapter to refer to the correct understanding of ‘is wrong’ in terms of relation R/Φ-ing.
understanding what John said within BAS. Her final option, I suggest, is to maintain that nihilism involves certain metanormative commitments that are incompatible with expressivism, and thus that she doesn’t have to be able to make sense of what John says. I argue, however, that John’s nihilism requires only one metanormative commitment—one to precisely the sort of objectivity that, I suggested above, seems incompatible with expressivism, despite the protestations of quasi-realists. Thus, in order for Jane to maintain that she does not have to be able to interpret John’s nihilistic claims, she must accept that her view is incompatible with such objectivity. And this, I believe, is far too large a bullet to bite.

3. Can John BE FOR Things?

If John, a nihilist, can coherently BE FOR things, then obviously BEING FOR does not itself entail making a SNJ. Here, it might seem the expressivist can just appeal to an analogy with what most cognitivists say about belief: Believing some things counts as making a SNJ, believing other things does not.

It’s not hard to see, though, why one might think that BEING FOR does involve making a SNJ in virtue of the kind of attitude that it is (in contrast with a cognitivist understanding of belief). Part of what’s supposed to be appealing about expressivism is that it vindicates our sense that SNJ has a distinctive connection with action. When I judge that something is right, wrong, good, bad, etc., it seems that I thereby commit (and perhaps am even motivated) to doing something (or not doing something). On a cognitivist view, this is surprising given wide acceptance of the idea that beliefs
alone can’t lead us to act; we need a desire, or something like it, for that.\textsuperscript{15} But if SNJs are more like desires anyway, the puzzle evaporates. Hence the appeal of expressivism.

Once we move to BAS, however, SNJs stop looking so distinctive. It turns out that all judgements involve the same non-cognitive attitude: BEING FOR. And BEING FOR, as Schroeder himself says, commits one to doing something. This indicates that, on the face of things at least, all judgements have the same link with action that substantive normative ones do. And so expressivism seems to lose its traditional impetus—its ability to make sense of how SNJs differ from judgements of other kinds.\textsuperscript{16}

This is not definitive, of course. An expressivist can maintain that, even though all BEINGS FOR commit one to doing something, only certain BEINGS FOR commit one in such a way, or to doing such particular things, that they count as SNJs. Personally, I think that only SNJs commit one to (not) doing things, and thus that BEING FOR (if there is any such attitude) entails making a SNJ.\textsuperscript{17}

But let’s suppose that I’m wrong. Let’s suppose that other kinds of judgement can commit one to (not) doing certain things, and thus that John can BE FOR anything at all. And that means that Jane can’t plausibly understand him as BEING FOR not Φ-ing murdering.

\textsuperscript{15} Of course, not everyone believes this, but among those who do, it is a large part of what speaks in favor of expressivism.

\textsuperscript{16} One way of putting the point is to say that BAS expressivists have the opposite problem that cognitivists do: Cognitivists say that certain beliefs are special in that they have a distinctive connection with action, and it is notoriously difficult to explain why. BAS expressivists, on the other hand, have to maintain that certain judgements are distinctively SNJs even though all judgements turn out to have a connection with action.

\textsuperscript{17} Or, more exactly, I think that only intentions commit one to (not) doing things, and that only SNJs commit one (in the relevant sense) to (not) intending things.
something about Φ-ing itself, or about the combination of BEING FOR and Φ-ing, that explains why
‘BEING FOR Φ-ing murdering’ is a SNJ while ‘BEING FOR not Φ-ing murdering’ is not.

It is hard to see, though, what Φ-ing could turn out to be. Most things—blaming for,
intending not to, being disposed not to—won’t fit the bill; being committed to bearing these
relations to murdering does not entail judging that murdering is wrong. Just as I can commit to
blaming for things I judge not to be wrong, so too can I commit to intending not to or being
disposed not to do them. We might even imagine this as a sort of analogue of Moore’s (1903) Open
Question Argument: “I know I’m committed to Φ-ing X, but do I think X is wrong?” seems open
for any relation I can think of (with the trivial exceptions of explicitly moral relations like ‘judging
wrong’).18

Now, to be fair, little has been said about what BEING FOR is like, so perhaps once we
understand what kind of commitment the expressivist is talking about, it will turn out that being
committed in this way to Φ-ing X (for some Φ) entails judging that X is wrong. But it’s at least as
hard to see what kind of commitment that could be as it is to see what Φ-ing could be.19

Let’s recap: Schroeder has offered the expressivist a recipe for developing a semantics that
avoids the negation problem (and, with further development, other aspects of EP). According to

18 Merli (2008) offers an argument that is similar to, though much more sophisticated than, the sort of Open
Question Argument I just alluded to—in the form of a Moral Twin Earth argument (following Horgan and
Timmons (e.g., 1991 and 1992)). Björnsson and McPherson (unpublished manuscript) propose a solution to
this problem. While I cannot hope to address this issue satisfactorily here, I do not think that this view can be
easily or plausibly adapted for use by the BAS expressivist to solve the problem at hand.

19 And, remember, the nature of Φ-ing itself must play a significant enough role in making BEING FOR Φ-ing
X a SNJ such that BEING FOR Φ-ing X is in some relevant way different from, say ‘BEING FOR proceeding as
if P’, which is apparently not a SNJ.
that recipe, SNJ involves BEING FOR—being committed to—bearing certain relations (the ones that constitute the semantic values of normative predicates) to the object of judgement. Thus, judging that murdering is wrong is ‘BEING FOR Φ-ing murdering’. And judging that murdering is not wrong is ‘BEING FOR not Φ-ing murdering’. If nihilists like John can coherently judge that murdering is not wrong, then expressivists need to be able to explain why ‘BEING FOR Φ-ing’ is a SNJ but ‘BEING FOR not Φ-ing’ is not. This raises two problems: First, it is natural to think that BEING FOR always involves SNJ, because BEING FOR seems to capture precisely the sort of action-guidance that, arguably, distinguishes SNJ from other kinds of judgement. (And again, this a distinction that expressivism is supposed to be uniquely able to explain precisely because it takes SNJ to involve a non-cognitive attitude—an advantage that is apparently lost when it turns out that the attitude in question undergirds all judgement.) Second, setting this worry aside, it is very hard to imagine what BEING FOR and the relations in question could turn out to be such that while BEING FOR alone does not involve making a SNJ, BEING FOR bearing these relations to things does, while BEING FOR bearing various other relations (e.g., not blaming for) does not. At the very least, the onus is on expressivists to give us some reason to think that they can construct a plausible theory in normative psychology to fill out the BAS schema.

Assuming I am right that Jane will not be able to understand John as BEING FOR not Φ-ing murdering while remaining consistent qua nihilist, it seems that if she is to make sense of what he said at all, Jane must either reinterpret John’s assertion, or deny that his assertion is consistent with his nihilism. Let us consider each response in turn.

4. Re-Interpretation

Earlier, we considered three different things that might be true of you when you judge that murdering is not wrong (continuing to generalize from blaming to Φ-ing):
4F. You are not FOR Φ-ing murdering.

5F. You are FOR not Φ-ing murdering.

6F. You are FOR Φ-ing not murdering.

Suppose that Jane accepts that BEING FOR itself entails making a SNJ. She might suggest that John is not to be understood in keeping with (5F), but rather with (4F) or (6F). Obviously, the latter won’t do (John does not think that not murdering is wrong). But it does sound quite plausible that, as a nihilist, John is not FOR Φ-ing murdering (after all, he doesn’t think that murdering is wrong).

Unfortunately, Jane is not merely trying to explain what’s going on in John’s head; she is trying to offer an account of what he expressed when he said that murdering is not wrong. So more needs to be said.

One worry about understanding John as expressing his not BEING FOR Φ-ing murdering is that it may not be possible to directly express the absence of an attitude at all. Arguably, when we express something we indicate what mental state(s) we are in. But while it is quite plausible that having an attitude is a particular mental state (or set thereof), it is not at all clear that lacking an attitude is.

Indeed, the expressivist may be committed to this line of thought even more than others. In other work, Schroeder (2008a) argues that expressivists should accept a view on which expression is understood in terms of claims’ “assertability conditions”: “Every sentence in the language is associated with conditions in which it is semantically correct to use that sentence assertorically” (Schroeder 2008a, 108). And those conditions, Schroeder argues further, must involve some mental state that the speaker is in. Thus, when I assert something, if we assume that I am speaking

---

20 As opposed to “indirectly” expressing it—reporting the absence or expressing other attitudes that entail or imply it. I discuss the possibility of appealing to such indirect expression to solve the problem shortly.
correctly—i.e., am speaking in keeping with the assertability conditions for the claim I make—it follows that I am in the relevant mental state. For example, if I say “the Earth is round,” arguably the assertability conditions for this include my believing that the Earth is round. And thus when I say this, I express that belief, which is a particular mental state. But again, it seems doubtful that there is any particular mental state (or set thereof) that is the absence of, say, anger. And so it is doubtful that not being angry can alone be a condition under which it is appropriate for me to assert anything.

In any case, even if it is possible to directly express the absence of an attitude, the negation problem remains unsolved. After all, Jane is not merely trying to say that John expressed something, but that he expressed something such that he disagrees with someone who thinks that murdering is wrong. But it’s not clear that this would be the case if John is merely expressing a lack of BEING FOR. If I believe that grass is green and you don’t (perhaps you’re colorblind), it is not at all clear that we disagree. In order to take this line, then, Jane would have to both show that she can understand John’s utterance as directly expressing the absence of BEING FOR and that, in apparent contrast to all other attitudes, BEING FOR something is sufficient for disagreement with those who are merely not FOR it.

At this point, Jane might look for help in the observation that we often indirectly express the absence of an attitude by directly expressing other attitudes that indicate or entail that absence. Indeed, given that, as just argued, the mere absence of an attitude isn’t enough to explain disagreement, an “indirect expression” model might be just what Jane needs to ensure that John can be understood to have some substantive mental content that’s in conflict with the judgement that murdering is wrong.

A typical way to “indirectly express” the absence of an attitude is to report it—to express the belief that one lacks it. Thus, “I don’t believe that the Earth is flat” expresses my belief that I don’t
believe that the Earth is flat. Assuming our beliefs are typically transparent to us, it is reasonable to conclude that by expressing this second-order belief, I accurately report the absence of a first-order belief in the Earth’s being flat. But this won’t help Jane, because the earlier problem remains: If I believe that I don’t believe that the Earth is flat, I don’t thereby disagree with someone who does believe that the Earth is flat.

Can one indicate one’s lack of an attitude in some other, non-reporting way? Consider the following case: I am sitting in my office, staring placidly at my computer, humming cheerfully to myself. You walk by and think to yourself, “David is not angry.” Did I somehow convey to you my lack of anger? Plausibly, yes. How did I do this? Arguably, in humming cheerfully to myself, I expressed certain other attitude(s)—call one SERENITY—that I do have. And SERENITY (let’s suppose) is generally, or perhaps even necessarily, incompatible with anger. Thus, when you see my expression of SERENITY, you reasonably infer that I am not angry.\(^{21}\)

Jane might suggest, then, that John’s utterance is similar to my humming to myself in my office; we each indicate the absence of an attitude—respectively, BEING FOR and anger—by expressing other attitudes we do have. In my case, we called the attitude SERENITY. Let us call John’s attitude NIHILO. Taking this tack, the expressivist proposes that when John says that murdering is not wrong, he is expressing NIHILO, and thereby indicating (through implication or entailment) his lack of BEING FOR.

Unfortunately, this proposal re-raises Schroeder’s concerns about unexplained inconsistency. Either NIHILO is defined in terms of BEING FOR (presumably as not BEING FOR) or it is not. If it is, then the expressivist is back where she started—trying to explain how to express a lack of BEING

\(^{21}\) My thanks to Mark Wells for this example, and for helpful discussion on the topic of the expression of absent attitudes.
FOR (either directly, in which case she has to explain why BEING FOR and not BEING FOR are relevantly inconsistent, or by appealing to yet another attitude and starting the process all over again). So, BEING FOR and NIHILO must be independent, primitive non-cognitive attitudes. And now, just as in the case of TOLERANCE and DISAPPROVAL, we are entitled to ask why BEING FOR and NIHILO are inconsistent. The expressivist apparently has no answer.

At this point, the expressivist might be tempted to look back once again to my anger example for help. I admitted there, after all, that SERENITY might be *psychologically* incompatible with anger. In that case, it seems we need no explanation for why anger and SERENITY are inconsistent in order to understand my humming as expressing a lack of anger. Rather, I express that lack by acting SERENE simply because it would not be possible for me to feel both anger and SERENITY. But while this might help explain how Jane could interpret John in this way, it would not help with the broader negation problem. The issue is that there is an important difference between two attitudes’ being *incompatible*—in the sense that one cannot be in both states at the same time—and their being *inconsistent*—in the sense that one is rationally criticizable for being in both at the same time. Of course, if they are psychologically incompatible, then no one could ever be inconsistent in the relevant sense. But we *can* have cases of interpersonal inconsistency—i.e., of disagreement. And notice that when I am angry and you are SERENE, we do not therefore disagree. So even if it were the case that there is a distinct attitude, NIHILO, that is psychologically incompatible with BEING FOR, this would not be enough to solve the negation problem, because it would not be enough to show why John disagrees with someone who thinks murdering is wrong.

Summing up: A natural alternative to interpreting John as BEING FOR not \(\Phi\)-ing murdering is to understand him simply as not BEING FOR \(\Phi\)-ing murdering. Unfortunately, this line faces serious problems, in that (a) arguably, John cannot directly express an absence of BEING FOR (i.e., cannot express it other than via expressing some further attitude), and (b) even if he can, or if there are
other attitudes (cognitive or non-cognitive) he might be expressing that would indicate that he lacks the relevant BEING FOR attitude, this would still not be enough to show why John disagrees with someone who has such an attitude.22

5. Incoherence

If John cannot merely be understood to be not FOR Φ-ing murdering, and ‘BEING FOR not Φ-ing murdering’ is a SNJ, an obvious out for Jane would be to simply maintain that when John says “murdering is not wrong,” he is making a SNJ, and is thus inconsistent qua nihilist. Indeed, Jane would not be the first to suggest that nihilists are inconsistent in this way. The position is typically argued for as follows, summarized by Jonas Olson:

According to the standard interpretation of Mackie’s error theory [or nihilism23] a first-order moral claim like ‘Torture is morally wrong’ is false. According to the law

---

22 There are, of course, other, more radical, reinterpretations. For example, the expressivist could hold that John is expressing his BEING FOR not BEING FOR anything. But like most radical reinterpretations, this has certain bizarre implications. For instance, for this to be a general solution to the problem, it would presumably have to turn out that all claims John makes to the effect that normative properties are not instantiated (“murdering is not wrong,” “pain is not bad,” “lying is not vicious”) are semantically equivalent. I also suspect that such radical reinterpretations would run up against other arguments made in this chapter, though I unfortunately do not have the space to explore this issue further here.

23 Error theory is a metanormative alternative to expressivism. Throughout this chapter, I have been using the term “nihilism” to refer to one commitment of error theory: that nothing is right, wrong, good, bad, etc. As I highlight in §6, it seems quite plausible that people make these kinds of first-order nihilistic claims without embracing error theory’s further metanormative commitments (such as cognitivism, which would clearly beg
of excluded middle it follows that its negation, ‘Torture is not morally wrong’, is true. That torture is not morally wrong would seem to imply that torture is morally permissible. More generally, then, the apparent upshot is that contrary to Mackie’s contention moral error theory does have first-order moral implications. And rather vulgar ones at that; if moral error theory is true, any action turns out to be morally permissible!

But it seems that we can also derive an opposite conclusion. According to moral error theory, ‘Torture is morally permissible’ is false. According to the law of excluded middle it follows that torture is not morally permissible, which seems to entail that torture is morally impermissible. More generally, then, the apparent upshot is that any action is morally impermissible! This may not be a vulgar first-order moral implication, but it is surely absurd. It also transpires that the standard formulation of moral error theory leads to a straightforward logical contradiction since we have derived that it is true that, e.g., torture is morally permissible (since any action is morally permissible) and that it is false that torture is morally permissible (since any action is morally impermissible). (Olson 2010, 68–69)

This argument trades on an unfortunate feature of moral (and normative) language, which is that “permissible” is sometimes understood to mean merely “not wrong,” and at other times understood to mean something normatively substantive, such as that a permissible action is an action that ought to be permitted. It is the very fact that these meanings are rarely distinguished that generates the

---

appearance that nihilism is incoherent. After all, if “permissible” means “ought be permitted” and “impermissible” means “ought not be permitted,” and everything is either permissible or impermissible, then every action either ought be or ought not be permitted. If that were the case, then of course nihilism would be incoherent.

Once we spell all this out, it’s not hard to develop a semantic view that avoids such conflation. Here are three: (a) reject the idea that permissible and impermissible are contradictories, allowing for some third category that is non-normative; (b) accept that “permissible” means nothing over and above “not wrong” and thus that permissibility is non-normative; (c) reject the idea that there is an entailment between “permissible” and “not wrong” and between “impermissible” and “wrong,” thus retaining permissible and impermissible as normative categories but leaving “not wrong” as a non-normative category.25

The problem, as we are now in a position to see, is that, arguably, BAS builds the conflation of “not wrong” and “ought be permitted” into normative semantics. When you say that something is wrong, you express BEING FOR Φ-ing it. And this means committing yourself to responding to it in certain ways—perhaps blaming for it, condemning it, avoiding doing it yourself, etc. When you say that same thing is not wrong, you express a commitment to not responding in those ways—to not blaming for it, to not condemning it, to not avoiding doing it yourself, etc. One way of seeing my arguments about BEING FOR seeming to involve SNJ is as pointing out that given what ‘BEING FOR

25 I prefer (b); Olson (2010) seems to pull for (c). I reject (c) because, like (a), it requires one to give up the idea that permissible and impermissible are contradictories, lest it turn out that all actions are either permissible or impermissible and thus that all actions fall into some normative category. So (c) doesn’t really solve the broader problem. I reject (a) because I think it is too difficult to accept that ‘impermissible’ doesn’t just mean ‘not permissible’. In any case, which solution we should accept is not the issue here.
not Φ-ing X’ seems to commit one to, it is apparently much more like judging that X ought be permitted than that it is merely not the case that X is wrong.

In any case, the upshot here is that holding that nihilists can’t coherently claim that murdering is not wrong means maintaining that though it seems quite obvious that we can distinguish two senses of “not wrong”—one substantively normative and the other not—we actually can’t, and thus nihilists can’t maintain that murdering is not wrong without making a SNJ. This seems far too large a bullet to bite.

Before moving on, it should be noted that there is a deeper kind of incoherence that the expressivist might appeal to instead. Again, if BEING FOR involves SNJ, then nihilists can’t BE FOR anything. So nihilists can’t even believe that nihilism is true. Indeed, this is a line that some cognitivists have pushed in terms of the normative nature of belief—if belief itself involves SNJ, then nihilists can’t have beliefs.26 But even this is not enough for the expressivist, because the arguments in the relevant literature are about why, given the nature of belief, we cannot really believe that nihilism is true. First, this does not show that nihilism is false, and thus that it is not a view that needs to be taken seriously. Indeed, it has even been argued that the fact that we are incapable of believing nihilism speaks in its favor.27 Second, and more germane, even if we can’t fully believe nihilism, it still seems possible to think that murdering is not wrong without making a SNJ about

26 See, especially, Matt Evans and Nishi Shah (forthcoming). For other arguments about the literal unbelievability of nihilism, see Streumer (forthcoming) and Suikkanen (forthcoming). Note that if I’m right that BEING FOR involves SNJ, it follows that expressivists are committed to (some version of) this controversial thesis about the nature of belief.

27 Streumer (forthcoming) argues that “our inability to believe the error theory undermines many objections that have been made to this theory.”
murdering (even if it necessarily involves a SNJ somewhere down the line, such as about the correctness conditions for one’s belief). And that is what expressivists have to deny.28

6. Making Room for Doubt

I have characterized our central character, John, as a normative nihilist. But, the expressivist might point out, most nihilists in philosophy are actually error theorists. Error theory is a metanormative view in opposition to expressivism. When an error theorist says “murdering is not wrong,” we may think that he is not really deploying first-order normative language at all, but is making a metanormative claim. In that case, arguably expressivism would not need to capture what John means when he says “murdering is not wrong” any more than it needs to capture what the non-naturalist means when he says “murdering has the Platonic property of goodness.”

But nihilism apparently extends outside of philosophical ethics—to people who have few or no metanormative commitments. Think about every person who has ever been struck by the possibility that life is meaningless, that nothing we do or fail to do matters, and thus that nothing is really right or wrong, good or bad at all. One need only search the Internet—or the shelves of a library—for phrases like “life is meaningless,” “existential crisis,” and “nihilism” to see just how far this phenomenon spreads outside of the discipline:

“Everything is permitted.” – Fyodor Dostoevsky, The Brothers Karamazov

“Are you feeding the poor? Are you Shakespeare? Doesn’t matter. It’s all insignificant in the end. If not today, then tomorrow. If not in a thousand years, then

---

28 Of course, none of this means that expressivists cannot offer an independent argument that nihilists really can’t coherently maintain that murdering is wrong. But I know of—and can imagine—no such argument.
in a million billion. The whirling clusters of galaxies don’t even notice.” – Jim Sloman, http://www.manyblessings.net/nothingmatters.html

“Nothing really matters, anyone can see.” – Queen, Bohemian Rhapsody

Indeed, I believe one needs only one (implicit or explicit) metanormative commitment in order to entertain nihilism: that normative thought and language are compatible with the possibility of objective normative truth. It is because our normative thought and language is such that we can think about what matters, as opposed to merely what matters to us, that it makes sense to wonder whether there is any fact of the matter at all, independently of wondering whether I know my own mind about it. To take a mundane example, you may have no doubts whatsoever that you are currently having a screen-like or paper-like visual experience (depending on how you’re reading this).

Nevertheless, surely you can wonder whether there really is a screen or a piece of paper in front of you. This is because our thinking about screens is appropriately objectivist—we take it that screen-facts and paper-facts might well concern things “out there” in the world. Analogously, it seems I can be absolutely certain what kinds of attitudes I (or anyone, or everyone, for that matter) have towards murdering and yet still wonder whether murdering is really right or wrong at all. And just as one can extend from wondering about the screen or paper to Cartesian doubt in the existence of the external world, one can extend from the murdering case to doubt about the rightness, wrongness, goodness or badness of anything—to a kind of nihilistic doubt.30

---

29 With the obvious exception of beings that are stipulated as having full normative information or the like.

30 The two are not precisely analogous. Cartesian doubt is not usually seen as doubt that there is anything “out there,” but rather just that what is out there might not be as it seems. In the normative case, though, nihilistic doubt is all-encompassing—one doesn’t merely doubt whether the normative facts are as one thinks they are; one doubts that there are any normative facts at all. If anything, though, this disanalogy strengthens my case, for while we can respond to the Cartesian skeptic by saying something like, “Well, I guess I can’t be sure that
7. Conclusion

According to BAS, judging that something is wrong means **BEING FOR** bearing relation \( R \) to it (\( \Phi \)-ing it), where **BEING FOR** is a general non-cognitive attitude of judgement and \( R \) is the semantic value of the predicate ‘is wrong’. In the same vein, BAS holds that judging that something is **not** wrong involves **BEING FOR** **not** \( \Phi \)-ing it.

Given the relationship between **BEING FOR** and action, I have suggested that **BEING FOR** is plausibly understood as involving substantive normative judgment—judgement that some normative property is instantiated.\(^{31}\) If so, it is a problem for this view that some judgements (especially, judgements that things are not wrong) are **not** SNJs. It follows that BAS expressivists must either (a) deny that **BEING FOR** (or at least **BEING FOR** **not** \( \Phi \)-ing something) always involves SNJ, (b) deny that all assertions that murdering is not wrong involve **BEING FOR**, (c) deny that nihilists can coherently maintain that murdering is not wrong (or, perhaps, maintain *anything*), or (d) deny that nihilists’ claims are ones that expressivists need to make sense of.

I have argued that the first three moves will be unsuccessful, or at least are likely to be. The last move (d) is to hold that nihilistic assertions carry with them metanormative commitments that beg the question against expressivism, and thus that expressivist semantics do not need to accommodate these claims (much in the way that, arguably, they do not need to accommodate claims about goodness’ being a Platonic entity). But, I maintain, nihilism’s only metanormative

---

\(^{31}\) Or, again, that some normative predicate is properly applied. See note 12 in this chapter.
commitment is the objectivist nature of normative thought and language. So expressivists can only make this move by denying that nature.

Given this, it seems to me that claiming that expressivism does not have to accommodate nihilism is like claiming that the Euthyphro Dilemma begs the question against Divine Command Theory with its commitment to the non-arbitrariness of the normative. Such a response would not get the Divine Command Theorist very far, not because it isn’t, strictly speaking, true, but because we are too confident in the non-arbitrariness of the normative for this response to carry any weight. In the same vein, I do not think denying that we can think and talk about objective normative truth is at all plausible. But doing so, I contend, is likely the only way for the expressivist to salvage her theory in the face of the negation problem.
CHAPTER TWO
CAN MOOREAN REALISM SHOULD ITS EXPLANATORY BURDEN?

G. E. Moore (1903) famously held that moral properties are “autonomous,” that they are distinct from properties of any non-moral, “natural” kind. Modern non-naturalists differ from Moore, and from one another, in a variety of ways. But centrally, they still maintain some version of this “autonomy” thesis, often with respect to normativity more generally, rather than just morality.¹ This thesis may be characterized in a number of ways. For the purposes of this discussion, I understand it as follows:

AUTONOMY. Normative properties are neither identical with, constitutive of, nor constituted by non-normative properties.²

¹ Some of the arguments from other theorists that I discuss in this chapter are made in terms of morality in particular, rather than normativity more generally. In order to avoid various caveats, parentheticals and orthogonal discussions, I will nevertheless uniformly speak of normative facts, properties, etc. Given that (a) the arguments I discuss can arguably all be extended to the normative more generally and (b) most modern non-naturalists are non-naturalists about normativity in general, I am confident that this slight fudge will serve only to simplify, rather than mislead.

² Enoch does not address the nature of constitution in his discussion of the epistemological challenge. However, earlier in the book, in his discussion of naturalism, Enoch appeals to a “nothing-over-and-above” relation: “I take the more philosophically-sounding relations of identity, reduction, constitution, and grounding to be attempts at precisifying this more intuitive relation . . . I take it to be a starting point for the debate over the relation between statues and ‘their’ lumps of clay that there is a sense in which the statue is nothing over and above the lump of clay” (Enoch 2011, 101–102). The apparent suggestion here is that to say that the normative is neither identical with, constitutive of, nor constituted by anything non-normative is to
Most self-proclaimed non-naturalists (along with others, perhaps\(^3\)) would assent to AUTONOMY. To avoid certain problems with the label “non-naturalist,” including the fact that some self-proclaimed non-naturalists might not accept AUTONOMY in this form, I will call those that do “Moorean realists.”\(^4\)

say that the normative is not related to anything non-normative in the way that statues are related to “their” lumps of clay (nor in the way that the clay is related to the statue). Ultimately, I don’t think that the “nothing-over-and-above” relation is much more helpful than the constitution one. After all, Enoch suggests that this relation is intuitive, but it then appears (given his inclusion of “a sense”) that even this relation needs to be disambiguated before it can do the work we need it to do. Nevertheless, I think the clay/statue analogy is fairly apt, and that the relevant relation—whatever it is best called—is intuitive enough that a more thorough theoretical discussion can be left off for another time.

\(^3\) Technically, there is no reason one couldn’t accept AUTONOMY but maintain that normative properties are in some sense “natural.” Historically, however, I know of no one who maintains such a view.

\(^4\) Well, the overwhelming majority of them, anyway. In the next section, I narrow the field slightly by introducing two further theses of Moorean realism, as I understand it. The list of Moorean realists—i.e., of theorists who, arguably, would endorse AUTONOMY and the other theses to be presented—includes Cuneo (2007), Dworkin (1996) and (2011), Enoch (2011), Fitzpatrick (2008), Huemer (2006), Kramer (2009), Nagel (1986) and (1997), Parfit (2011), and Scanlon (2009). There are at least three notable absences here. The first is Oddie (2009), who I set aside because he holds the minority view that the non-natural normative properties are causally efficacious. I discuss this issue further below. Second is Shafer-Landau (2005). Though he considers himself a non-naturalist, his claim that natural facts “exhaustively compose moral ones” is apparently at odds with AUTONOMY (Shafer-Landau 2005, 75). Finally, there is Wedgwood (2007). On the one hand, he is clear that the normative can be understood in terms of the mental (and not in terms of anything else), which seems clearly to violate AUTONOMY. On the other hand, he believes that the mental cannot be understood without an appeal to the normative—and so there is a sense in which we might
Standard complaints about Moorean realism centrally include its inability to plausibly explain our epistemic access to the “queer” normative facts the view posits, the metaphysical relationship between the normative and the non-normative (e.g., why the former supervenes on the latter), and how our normative terms succeed in referring to normative properties. In this chapter, I set out to assess the Moorean’s ability to shoulder these burdens. In what follows, I discuss each of these challenges, and the Moorean’s ability to explain the relevant phenomena. I conclude that while the Moorean may not find the combined force of these objections utterly devastating, the burden placed on him is extraordinarily heavy, heavy enough that we should have serious reservations about accepting a Moorean metanormative theory.

1. Epistemology

1.1 The Challenge

Epistemological challenges to Moorean realism take many forms. Recently, the major focus has been Sharon Street’s (2006) “evolutionary argument.” In short, Street contends that we should reject Moorean realism because our normative beliefs are explained by selective (evolutionary) pressures.

Street’s argument is unique in some respects, but it shares important features with other arguments in metanormative theory. Compare, for example, John Mackie’s (1977) “argument from attribute to him a kind of “wide Moorean” view, holding that the normative/mental realm is autonomous (though, like Oddie, Wedgwood also holds that the normative is causally efficacious). I discuss Wedgwood further in §3, where I address normative semantics.

5 For the original “queerness” challenge, see Mackie (1977).

6 See, e.g., Blackburn (1973) and (1988b).

7 See, e.g., Wedgwood (2007).

8 Street attacks a broader class than Moorean realism, but it is one of her targets.
relativity.” Mackie argues that acceptance of error theory may be justified by the fact that our normative beliefs are, at least for the most part, merely reflections of social practices.

Of course, Street and Mackie have very different views. Mackie believes that normative thought and language concern realistically-construed facts, but holds that the best explanation for our beliefs supports error theory.9 Street, on the other hand, denies that normative thought and language are thus concerned, holding instead that the correct explanation of our normative beliefs—an explanation itself in competition with Mackie’s—vindicates constructivism over realism.

Despite these differences, both arguments are instances of what Russ Shafer-Landau (unpublished manuscript) calls “genealogical arguments” against realism. The real force of these arguments comes not from their claims about what the explanation of normative beliefs is, but rather what it is not. If normative beliefs are not explained by realistically-construed normative facts, then any robust connection between our beliefs and those facts might seem miraculous.10 For Street, the mere fact that selective pressures explain our normative beliefs is not an objection to realism; it is the fact that that explanation seems to rule out an explanation in terms of normative facts. Likewise for Mackie, it is the fact that our normative practices are mere practices—i.e., that they reflect no further “objective” truth—that does the real argumentative work.

Genealogical challenges like these invoke (either implicitly or explicitly) a broader epistemological challenge to explain our epistemic access to realistically-construed normative facts. They do so by proposing a debunking explanation that would seem to undermine the realist’s ability

9 Mackie was an error theorist about morality in particular but not about normativity more generally. However, as has been noted by others, some of his arguments seem to extend naturally to the normative. See, e.g., Bedke (2010).

10 Of course, our beliefs could instead explain the facts; but that wouldn’t be a very “realist” thing to say—more on this later.
to meet that broader challenge. But the challenge can be made more direct, as it is, for example, in Mackie’s (1977) other central argument for error theory, the “argument from queerness.” Mackie contends that because they are metaphysically “queer,” none of our ways of knowing things could get us access to the normative facts. The precise nature of Mackie’s “queerness” worry is not entirely clear, but it hints at the fact that the epistemological challenge is especially troubling for those who embrace the idea that normative facts are unlike facts of other kinds—i.e., relevantly, for Mooreans. Other realists may have less difficulty. A naturalist, for instance, might just hold that our epistemic access to the normative facts is parasitic on our access to the natural properties normative properties reduce to.

Recognizing the diversity in expressions of the epistemological challenge, David Enoch (2011, chap. 7) offers a helpful formulation of the challenge, in the hopes of making it as strong as possible before addressing it on behalf of the Moorean. Whenever we take ourselves to experience some degree of epistemic success in a domain (i.e., take it that our beliefs in that domain are, to a significant extent, true, or justified, or reliably formed, or knowledge), Enoch argues, we are committed to a correlation between our beliefs and the facts in that domain. For instance, if we take ourselves generally to know or have justified beliefs about the weather, it seems we commit to a

---

11 Mackie’s argument is not just epistemological; it is supposed to be a direct metaphysical challenge as well. For our purposes, however, the epistemological aspects of the argument are most relevant.

12 This can plausibly be seen as a central case of “one man’s modus ponens…” The “queerness” that Mackie takes to vitiate realism seems intimately related to, if not the same as, the thesis of AUTONOMY that is so central for the Moorean. This may explain why many Mooreans see error theory as their “second best option,” for it seems to them that only the error theorist—but not the constructivist, the naturalist, the expressivist, etc.—takes normativity “seriously” in the way that they do. See, e.g., Enoch (2011), 115.

13 Enoch’s formulation of the epistemological challenge was originally offered in his (2010a).
correlation between our weather-beliefs and the weather-facts. Especially for Mooreans, given their commitment to the distinctive nature of the normative, the analogous correlation between normative beliefs and normative facts is, as Enoch puts it, “striking”; “absent some . . . explanation, the correlation would be just too miraculous to believe” (Enoch 2011, 158). In other words, the correlation seems unlikely to be brute, unexplainable. We can therefore understand the epistemological challenge as a challenge to explain the correlation between normative beliefs and normative facts—a correlation the Moorean is committed to if he is to avoid skepticism.

In order to offer such an explanation, the Moorean would be wise to consider how correlations are explained in general. Enoch offers:

[I]f the correlated factors are A and B, then (roughly speaking) either A-facts are somehow (causally, constitutively, or both) responsible for the B-facts, or the B-facts are responsible for the A-facts. . . . [Alternatively,] it is possible that the explanation . . . is in terms of a third factor, C, that is (roughly speaking) responsible both for A-facts and for B-facts. (Enoch 2011, 167)

I take it that by “responsible for” Enoch means something like “(part of) the ultimate explanation for.” Throughout, I use “responsible for” in this way. Further, I assume, as Enoch does, that the relevant forms of explanation are either causal or constitutive. Given this, it seems that the possible responses to the challenge are: (a) normative beliefs are responsible for (i.e., are (part of) the ultimate causal or constitutive explanation for) normative facts; (b) normative facts are responsible for normative beliefs; or (c) some third factor is responsible for both.

---

14 Any correlation between beliefs and facts seems “striking” in the absence of explanation but, as evidenced by the need for discussion on this very topic, the explanation is significantly less transparent in the normative case (in contrast with others, such as our beliefs about the current weather).
The Moorean’s options are more limited. The first option, (a), is ruled out not only for the Moorean, but for most (if not all, depending on your favored taxonomy) realists, because it violates the commonly-held thesis of mind/attitude/response/stance-independence:

INDEPENDENCE. Normative facts are not made true by their “ratification” from within any actual or hypothetical perspective.\(^{15}\)

All Mooreans endorse INDEPENDENCE. First, it is almost certainly entailed by AUTONOMY.\(^{16}\) Second, non-naturalists tend to be placed into the broader “objectivist” camp that is, arguably, defined partly by its acceptance of INDEPENDENCE, or something quite like it. In any case, INDEPENDENCE is a thesis of Moorean realism as understood here.\(^{17}\)

Returning to option (a): If my belief that eating meat is wrong were responsible for the wrongness of eating meat, then clearly eating meat’s wrongness would be made true by its

\(^{15}\) Different versions of this thesis have been called different things, but seem to be capturing roughly the same idea. This formulation is borrowed from Shafer-Landau’s (2005, 15–16) characterization of “stance-independence.”

\(^{16}\) The only way this could not be the case would be if some aspect of our perspective were normative. Given increasing acceptance of the view that belief has a normative component, I will not say that this is impossible. But I am confident that this view will not have any substantive effect on the strategies open to the Moorean. In any case, it is hard to imagine that any non-naturalist would deny INDEPENDENCE, either because they embrace a normatively-loaded conception of belief or for any other reason.

\(^{17}\) As indicated in the previous note, it’s unclear that we really need to talk about INDEPENDENCE as an independent thesis. I do so largely because I am following Enoch’s lead, who relies on “response-independence” (rather than the broader AUTONOMY) in rejecting certain kinds of explanation of the correlation between normative beliefs and normative facts.
“ratification” from within my perspective. This would violate INDEPENDENCE. Thus, as Enoch himself concludes, Mooreans cannot offer solutions of form (a).18

Turn now to option (b), on which the order of explanation is reversed—the ultimate explanation of our normative beliefs is the normative facts. Again, I assume, with Enoch, that such an explanation would be either causal or constitutive. AUTONOMY straightforwardly rules out the latter, since it holds that there can be no constitutive relationship between the normative and the non-normative. So if the normative facts explain our normative beliefs, they must do so causally. Unfortunately, it is now widely accepted that the autonomous normative properties the Moorean

18 INDEPENDENCE also rules out some instances of (c). According to “projectivists,” normative facts are ultimately explained by certain non-cognitive pro- and con-attitudes. For them, a third factor—the aforementioned non-cognitive attitudes—might explain the correlation between normative beliefs and normative facts because those attitudes are responsible for both the beliefs and the facts. Yet if the ultimate explanation for the wrongness of eating meat were my con-attitude towards eating meat, this would violate INDEPENDENCE just as surely as if the explanation were my belief. So Mooreans cannot employ this strategy, either.

Of course, certain forms of non-cognitivism also endorse the claim that normative facts are ultimately explained by certain non-cognitive pro- and con-attitudes. Such views are not subject to Enoch’s challenge, since it is explicitly about normative beliefs. However, some non-cognitivists—notably, quasi-realists—arguably face an analogous challenge. See Street (2011).

It should also be noted that while throughout I refer to normative “facts,” I do so only for the sake of consistency (as this is the phrase Enoch uses). For projectivists or quasi-realists, it might turn out (depending further still on their metaphysical views) that there are normative “truths” but no normative “facts.” This distinction is irrelevant to my arguments.
believes in would have to be causally inefficacious. Indeed, self-proclaimed non-naturalists themselves nearly all accept:

INEFFICACY. The basic normative facts are causally inefficacious.\(^{19}\)

This is the only thesis discussed so far that is likely to be at all contentious among non-naturalists.\(^ {20}\) While many believe that non-naturalism is committed to INEFFICACY, I cannot fully address this issue here.\(^ {21}\) So let us simply presume that the prevailing opinion is correct. From here out, all mentions of “Moorean realism” should be understood as referring to any view committed to AUTONOMY, INDEPENDENCE and INEFFICACY.

It follows from the above that the Moorean cannot embrace option (b), as tenets of his view—INEFFICACY and AUTONOMY—respectively rule out ultimate causal or constitutive explanations of the normative in terms of the non-normative. Thus, for our purposes at least, it would seem the Moorean is left with only one form of explanation for the correlation between normative beliefs and normative facts: He must maintain (c) that some third factor (that is not a pro-

\(^{19}\) Enoch explicitly endorses this claim. He does not, however, say what he means by “basic.” I take it the idea is something like this: Utilitarians believe that the fact (if it is a fact) that utility should be maximized is a (the only) basic normative fact. Another normative fact might be that I ought to perform a particular action right now (because that act would maximize utility). But this is not a basic normative fact because it depends, in part, on contingent, non-normative facts about my circumstances and the consequences of my actions. From here, when speaking of normative facts, I mean always to be referring to the “basic” ones thus understood.

\(^{20}\) For an attempt to reconcile non-naturalism with causal efficacy, see, e.g., Oddie (2009). This is a rare view, however. Most non-naturalists, including those mentioned in note 4 in this chapter (except for Wedgwood and Oddie himself), seem to accept INEFFICACY.

\(^{21}\) For interesting arguments against normative (moral) causation, see Slutsky (2001) and Zangwill (2006).
or con-attitude) is responsible for both normative beliefs and normative facts. Indeed, this is precisely the sort of solution Enoch claims to offer. Let us consider his proposal.

1.2 Enoch’s Proposed Solution

Ironically, Enoch’s solution begins with the premise that our normative beliefs are (largely) explained by selective (evolutionary) pressures—precisely the premise that Street takes to undermine realism. Street’s mistake, claims Enoch, is limiting herself to explanations of form (a) and (b). Once we recognize the possibility of (c), evolved normative truth-tracking becomes significantly less mysterious. This is because once one acknowledges the possibility of a third-factor explanation, it becomes apparent that truth-tracking might be guaranteed through a sort of “pre-established harmony” between normative beliefs and normative facts (Enoch 2011, 168). And this would be the case, Enoch argues, given the minimal and plausible assumption that survival is (at least sometimes) good.

Here’s how this works: Suppose that survival is good. Evolution typically selects for features that promote survival. Thus things explained by selective pressures by-and-large promote survival. Since survival is good (we are supposing), it follows that those things explained by selective pressures will (happen to) by-and-large promote the good. So our normative beliefs, if explained by selective pressures, by-and-large promote the good. Finally, if we assume that we best promote the good when our normative beliefs are true, then it will turn out that selective pressures have given us many true normative beliefs—enough, presumably, to establish the correlation in question.

22 Or, at least, “a close relative of truth-tracking,” if one takes tracking to entail a causal connection. See Enoch (2011), 166 n. 36.

23 This is plausible but not uncontroversial. Some utilitarians, for example, believe that while the only basic normative fact is that utility is the good, it may be better that people believe some “folk” normative view that leads them to promote utility than for them to believe utilitarianism itself, which would lead them only to
If we grant all this (and that’s a big “if”), it seems that Enoch has met his own challenge. He has offered a third-factor explanation of the correlation between normative beliefs and normative facts. That third factor is the goodness of survival.

But there is a potential problem here. According to Enoch, the correlation between normative beliefs and normative facts might be explained in one of three ways, two of which are ruled out by his metanormative commitments. The remaining possibility is (c) that some third factor is responsible for both normative beliefs and normative facts, and thus explains the correlation between them. Enoch proposes such a third factor: the goodness of survival. But notice that this “third factor” is no such thing, for it is itself a normative fact. Enoch’s solution, purportedly of form (c), seems to collapse into a solution of form (b). Rather than suggesting that the normative facts are generally responsible for normative beliefs, Enoch has instead suggested that a single normative fact—the goodness of survival—is responsible both for normative beliefs and (the rest of the) normative facts. Since Mooreans are committed to the view that normative facts cannot be responsible for normative beliefs, Enoch’s solution apparently falls prey to his own arguments.

Importantly, this is a problem not just for Enoch’s particular solution, but for any solution of form (c). Recall that according to Enoch, there are three forms of response to the epistemological challenge: (a) normative beliefs are responsible for normative facts; (b) normative facts are responsible for normative beliefs; or (c) some third factor is responsible for both. Because of

waste time making difficult and costly utility calculuses. See, e.g., Sidgwick (1874), 487-489. Enoch suggests, however, that there are additional reasons to find his view plausible, stemming from something like Gibbard’s (1990, chap. 4) “normative governance” mechanism. The idea, roughly, is that we are the sorts of creatures that act, much of the time, in accordance with how we believe we should act. So it makes sense that in order to “push” us to Φ, selective pressures would lead us to believe that we ought to Φ. This is also discussed at some length in Copp (2008), drawing on Kitcher (2005).
INDEPENDENCE, the Moorean cannot offer a solution of form (a). Because of AUTONOMY and INEFFICACY, the Moorean cannot offer a solution of form (b).

Now, any third factor would have to be either normative or non-normative. As we saw in the case of Enoch’s own solution, the third factor cannot be normative, because the normative can neither cause (because of INEFFICACY) nor constitute (because of AUTONOMY)—and thus cannot be the ultimate explanation for—anything non-normative.

The alternative is to look for a non-normative third factor. AUTONOMY rules out the possibility that anything non-normative constitutes anything normative. So the Moorean would have to appeal to a causal explanation. But this is a non-starter. Aside from its being manifestly bizarre to speak of causing happiness to be good or causing treating agents as mere means to be wrong, the fact that the basic normative facts are necessary (as most believe) seems to preclude their being causally explained at all. After all, causes are events within a world and thus cannot explain truths that hold across possible worlds. So this option seems closed as well.

It is imperative that we keep in mind here that we are talking about ultimate explanations. Like most everyone, Mooreans maintain that the normative supervenes on the non-normative, and thus hold that it is often the case that something’s non-normative properties are the immediate

---

24 My thanks to Christian Coons for this way of putting the point. The obvious exception to all this, of course, is Divine Command Theory, according to which God creates (causes?) the normative truth. I am confident that such theories are subject to sufficient objections along other lines (e.g., the Euthyphro dilemma) to be set aside here. In any case, such views are typically referred to as super-naturalist rather than non-naturalist, and are not particularly friendly to Moorean intuitions.
explanation for its normative properties. But for the Moorean, at least, such cases must be representative of some further basic normative fact that is part of the ultimate explanation in these cases. For example, if Bob is bad because Bob is a liar, it seems Bob’s being a liar explains (in the immediate sense) his being bad. But this is true, for the Moorean, only because (say) it is an independent normative fact that being a liar makes one bad. Ultimately, Bob’s badness depends not just on his being a liar, but also on that normative fact.

There is thus a general problem for the Moorean: On his view, the normative cannot be the ultimate explanation for the non-normative, nor can the non-normative be the ultimate explanation for the normative. It follows that the Moorean cannot offer solutions to the epistemological challenge of any of the three forms Enoch has proposed. So if Enoch is right that the challenge can only be met with a solution of form (a), (b) or (c), then the Moorean cannot meet the challenge at all.

Luckily for the Moorean, Enoch is wrong to think that he is thus limited. For not only is there a fourth form of solution available, but, as I argue below, Enoch’s proposal is of this fourth kind—he misidentifies the form of his own solution.

1.3 The Form of Enoch’s Solution

The worry that Enoch’s solution collapses into a solution of form (b) seems natural given a strict reading of Enoch’s discussion of the forms a solution might take. After all, if normative beliefs and normative facts are not responsible for one another, then, Enoch has claimed, some third factor

---

25 I could, of course, just introduce a SUPERVENIENCE thesis as a fourth commitment of Moorean realism. But I take it that a commitment to supervenience of some kind is universal enough that this can remain implicit.
must be responsible for both—the solution must be of form (c). But Enoch’s solution is not, strictly speaking, of this form at all. To see what form his solution does take, consider Enoch’s example of how third-factor explanations function in other contexts:

What explains the correlation (if indeed there is one) between giving rise to strongly affectionate feelings and having a poor sense of time? Here’s one possible answer: Young children are cute. Being a young child explains—indeed, perhaps causes—having a poor sense of time. And of course, being cute is closely though perhaps not causally related to giving rise to strongly affectionate feelings. The fact that young children are cute, then, pre-establishes the harmony between giving rise to strongly affectionate feelings and having a poor sense of time. (Enoch 2011, 169)

The factors correlated are (A) giving rise to strongly affectionate feelings and (B) having a poor sense of time. What explains this correlation, according to Enoch, is (C) the fact that young children are cute. What is key here is that we notice that the fact that young children are cute is itself responsible for neither A nor B.

If C is not responsible for A or B, how can it explain their correlation? The answer is not hard to find: Cute things tend to garner affection. Young children tend to have poor senses of time. And in cases exemplifying the correlation between A and B, certain things’ being cute explains their garnering affection and those same things’ being young children explains their having a poor sense of time. So there is a correlation between A and B not because something is responsible for both, but because what (in some cases) explains some particular thing’s instantiating A is correlated with what (in some of those same cases) explains that same thing’s instantiating B. This further
correlation—between being cute and being a young child—has simply been expressed, by Enoch, as the single claim (C) that children are cute.26

Here is what we have discovered: The most obvious way for a correlation to be explained is through one correlate’s being responsible for the other. A less obvious way, as Enoch tells us, is for some third factor to be responsible for each correlate. But yet another possibility is that the target correlation (i.e., the one we are attempting to explain) between A and B is explained by another correlation whose correlates are respectively responsible for A and B.27 It is an explanation of this last kind—what I will refer to as form (d)—that is at work in Enoch’s example above: The correlation between being cute and being a young child explains the correlation between giving rise to strongly affectionate feelings and having a poor sense of time.

Return now to Enoch’s solution: Enoch proposes that survival’s goodness is the third-factor explanation of the correlation between normative beliefs and normative facts. Since survival’s goodness is a normative fact (if it is a fact), if it were responsible for both normative beliefs and normative facts this would indeed collapse into a solution of form (b). But this is not the case. After

26 It is clear that this expresses a mere correlation (as opposed to a necessary connection between properties) given that (I trust we’d all agree) not all children are cute.

27 The actual possibility is broader than this. It could turn out that, in some cases, the target correlates are not ultimately explained by the third-factor correlates but by something “further up the chain.” This might happen, first, if the target correlates are ultimately explained by the correlates of some nth-level correlation that in turn explains the n-1th level correlation, which then explains the n-2th level correlation, and so on. Alternatively, but similarly, it might be that the target correlates are ultimately explained by one of the correlates of some nth-level correlation in which that correlate explains its correlative, which in turn explains the n-1th level correlation and so one. I do not explore these possibilities further here, as they should have no substantive effect on my arguments.
all, survival’s *goodness* is not responsible for our normative beliefs; rather survival itself (or, more properly, selective pressures that “aim” at survival) is.  

The situation here is similar to the non-normative case above: Enoch appeals in his solution to the goodness of survival. “Survival is good” expresses a correlation between survival and goodness—presumably, something like “things that promote survival tend to also be good.” This third-factor correlation explains the correlation between normative beliefs and normative facts because its correlates are responsible, respectively, for normative beliefs and normative facts. On this model, the fact that they promote survival is responsible for our normative beliefs being as they are.

---

28 Similarly, it is not *survival’s* goodness that is responsible for the (rest of the) normative facts; rather, it may be that the remaining normative facts follow in some way from the nature of goodness itself. Enoch offers: “Perhaps the evolutionary ‘aim’—whatever exactly it is—causally shapes our normative beliefs, but the fact that the evolutionary ‘aim’ is of value does not causally shape the normative truths. It seems more appropriate to say that it is related to them in some constitutive way, a way the details of which depend on the details of your favorite first-order, normative theory” (Enoch 2011, 169). It is worth noting, in fairness to Enoch, that given what he says here, as well as the fact that he consistently qualifies his remarks by talking of third factors’ being responsible for correlates only “roughly speaking,” it is plausible that he implicitly recognizes form (d). Nevertheless, it is important that we explicitly distinguish this possibility from form (c) so as to better understand how each form functions and, ultimately, better understand how Enoch’s purported solution is meant to work.

29 In the non-normative case above it was fairly clear what was correlated: the two properties of ‘garnering affection’ and ‘having a poor sense of time’. Here, it is less clear. Is it the case that, as just suggested, the correlation is between the properties of ‘promoting survival’ and ‘being good’? I’m genuinely uncertain. But it certainly seems as though *some* correlation is being expressed. I can think of no alternatives other than a property identity, or a causal or constitutive relation, all of which are ruled out by tenets of Moorean realism.
are. And the nature of goodness is responsible (in some presumably constitutive sense, Enoch suggests\(^30\)) for the other normative facts.

Summing up: Because Enoch claims that correlations are explained in one of three ways, and seemingly “shoehorns” his solution into one of the three, his solution might appear to collapse unacceptably into a form he rejects. But the problem is not that his solution thus collapses, but rather that it is of a fourth form Enoch does not discuss. And thankfully so, given that without a fourth option, the Moorean position would apparently be indefensible.

1.4 Other Proposed Solutions in the Literature

Enoch is not the only realist to respond to Street’s formulation of the epistemological challenge. Shafer-Landau (unpublished manuscript) identifies several forms of solution (beyond rejecting the evolutionary story itself or undermining particular versions of the challenge). The first relies on normative reduction and can therefore be set aside here, as such reduction is incompatible with Moorean realism.\(^31\) The second solution is to appeal to the a priori nature of normative truth. But saying that normative truth is a priori tells us very little about how the broader epistemological challenge is to be met. We are still entitled to ask for an explanation of the correlation between our beliefs and those a priori truths. And everything that has been said thus far to rule out various kinds of explanation for that correlation still applies. Thus, if my arguments above that only option (d)

---

\(^{30}\) See note 28 in this chapter.

\(^{31}\) For discussion of Street’s argument from the perspective of reductive realism, see Copp (2008).
remains are successful, maintaining that normative knowledge is a priori adds very little (if anything) to the discussion.32

The last form of solution is what Shafer-Landau calls the “indirect tracking hypothesis.” The name alone should make it clear that this refers to solutions like Enoch’s. But Enoch is not alone. Knut Skarsaune (2011) argues that if pleasure is good and pain bad, those who believe as much are more likely to avoid pain and seek pleasure, and those who do so are more likely to pass on their genes. Similarly, Erik Weilenberg (2010) suggests that we might have moral rights grounded in the very cognitive faculties that have been selected for in human beings.

Both Skarsaune and Weilenberg are responding to Street’s version of the epistemological challenge, not Enoch’s. But it is not hard to see that were they to adapt their solutions to address Enoch’s formulation, their solutions would look much like his own. In each case, some normative third factor—itself a correlation between the normative and the non-normative—would be responsible for the correlation between normative beliefs and normative facts. This convergence is further evidence that Moorean solutions to the epistemological challenge can take only the form I have suggested. Which particular response the Moorean should accept—Enoch’s, Skarsaune’s, Weilenberg’s, or something structurally similar but differing in the details—will be a matter of the plausibility of the particular normative fact(s) and evolutionary story the response invokes.

Here is where the Moorean stands: He has been challenged to explain a correlation between normative beliefs and normative facts. He can do so only by appealing to a further relation between the normative and the non-normative. Unless he is to maintain that there is an infinite regress of

---
32 For relevant discussion of the problems for regarding “intuitionism” (which has clear ties to the synthetic a priori and self-evidence) as a non-naturalist-friendly epistemology, see Bedke (2009). Enoch (2011, 163 n. 23) mentions Bedke’s argument as among those that hint at his formulation of the epistemological challenge.
such relations, at some point he must stop. And so at some level, there will be a relation between the
normative and the non-normative—a basic, substantive normative fact—that the Moorean cannot
explain.

Of course, the Moorean might deny that this final, unexplainable explanans must itself be a
correlation between normative and non-normative factors. Enoch (2011, 144), for example, seems to
think that survival’s goodness would be explained by the content of some norm. But it is clear that
whatever explanation the Moorean offers, at some level he will have to stop and admit that there is a
relation of some kind between normative and non-normative properties that cannot be explained.
And so it may be that even if such a brute fact can be appealed to to meet the epistemological
challenge, a metaphysical one remains. Simon Blackburn (1973 and 1988b), for example, famously
objects that non-naturalists cannot explain why the normative supervenes on the non-normative.
And this supervenience is, of course, a relation between distinct properties (for the Moorean, at
least).33

Indeed, perhaps the epistemological challenge is really just a special case of this broader
metaphysical challenge. One might plausibly understand Blackburn as pointing out that the Moorean is
incapable of explaining—all the way down—relations between the normative and the non-
normative. If we understand the epistemological challenge as Enoch has, and Blackburn’s challenge
as just described, then it seems that there is a general challenge to explain relations between the
normative and the non-normative, and that the epistemological challenge is a special case of this

33 As with “facts” vs. “truths” (see note 18 in this chapter), I ignore here various issues concerning the nature
of properties. On some views, what I’m calling something’s properties might just be something like “properly
applied predicates.” This should have no substantive impact on my arguments.
challenge, demanding an explanation for the particular (cor)relation between normative beliefs and normative facts.

Of course, we do not want to push this analogy too far: Obviously relations between beliefs and facts are not the same as relations between normative properties and the non-normative properties they supervene on. For one thing, we are talking about an immense number of correlations between each person’s beliefs and the facts, not just between two properties. Nevertheless, this poses an important dialectical challenge for the Moorean—at least for those, like Enoch, who take the epistemological challenge seriously: Given that he accepts that it would be a problem if he were incapable of explaining the relation in the epistemic case, Enoch owes us an explanation for why he may remain comfortable positing other relations between the normative and the non-normative that he cannot explain.  

In any case, somewhere down the line, the Moorean will have to appeal to a relation between the normative and the non-normative that he cannot explain. As I have just argued, the Moorean must either deny the need to meet the epistemological challenge in the first place or explain why it is that even though correlations between beliefs and facts cannot be brute, other relations between the

---

34 I have discussed this issue with Enoch. Given that we are trading in “plausibility points,” he argues, it seems that the Moorean loses fewer points in positing a single brute correlation (between survival and goodness) than he does positing many brute correlations between our many normative beliefs and the normative facts. He thus maintains that his argument makes progress against the epistemological challenge on behalf of Moorean realism. It seems to me, though, that the sheer number of correlations is irrelevant. For example, even if a monistic view is more plausible than a pluralistic one, surely it is not so just because it happens to posit fewer basic normative facts and thus fewer connections between the normative and the non-normative at the fundamental level. This leads me to suspect that one does not gain plausibility points (at least not in any significant quantity) simply by lowering the number of brute correlations posited by one’s view.
normative and the non-normative can. Obviously, there are those—like, apparently, Blackburn—who believe that no such explanation is possible, and thus that the Moorean’s inability to explain all relations between the normative and the non-normative is damning. Let us attempt to adjudicate.

2. Metaphysics

Simon Blackburn (1973 and 1988b) charges that the Moorean cannot explain normative supervenience. Above, I suggested that we can reframe this in terms of a broader version of Enoch’s epistemological challenge—a challenge to explain relations between the normative and the non-normative. But this is not necessarily a “challenge” in the sense that the epistemological one is, for many Mooreans hold the commitment to brute, basic normative truths to be a feature of their view, rather than a bug. Some, like Parfit (2011), do so explicitly and without hesitation. Others, like Enoch (2011), seem to feel a greater need to address the issue. But whatever their attitude, the above discussion serves, in part, to show that the Moorean really is thus committed: He cannot explain relations between the normative and the non-normative all the way down. So our first question should not be whether Mooreans can explain normative supervenience, but rather whether the inability to do so is a bullet that’s too big to bite (or a bullet at all, for that matter).

As Enoch points out, part of what makes this seem to be a bullet is that many people accept (some version of) *Hume’s Dictum* that there are no “necessary relations between distinct existences”

---

35 As in Street’s case, Blackburn’s target is broader, but includes the Moorean.
For the Moorean, the idea that the normative and the non-normative are distinct existences is central—arguably, it is just another way of expressing the thought behind AUTONOMY.

Enoch (2011, 147) thinks that the cost (in plausibility points) of rejecting Hume’s Dictum is “completely affordable.” He does admit, however, that the objection can be improved by weakening the principle. Tristram McPherson (forthcoming), for example, offers the more modest:

Modest Humean (MH). “Commitment to brute necessary connections between distinct properties counts significantly against a view.”

MH, unlike Hume’s Dictum, joins Enoch in making the issue one of “plausibility points” rather than of an absolute metaphysical “deal-breaker.” Its other important contribution is the inclusion of bruteness. This is important because the typical counterexamples to Hume’s Dictum—things like the necessary relation between triangularity and trilaterality or between Socrates and the singleton set \{Socrates\}—are not clearly brute; arguably we have something to say about why these properties always go together, such as concerning the nature of the set-membership relation (Enoch 2011, 147). It is much less plausible, it seems, to claim that certain distinct properties necessarily co-instantiate for no reason whatsoever. This lends MH intuitive force that the original Dictum lacks.

Though he admits all this, Enoch does not think that MH is a very good reason to reject his Moorean view. He concludes:

“I accept some brute . . . relation here between distinct existences, and so that [Moorean] Realism stand [sic.] in violation of Hume’s Dictum. It is this result that, I argue, is not intuitively damaging. And if what remains of the venerable

---

36 See Hume (1739), Book I, Part III, Section VI.

37 Though those like Parfit (2011) who are committed to a “non-ontological” version of Moorean realism might object to the term “existences.”

38 Enoch (2011, 147) quotes here from a draft version of McPherson (forthcoming).
supervenience challenge . . . is just the need to reject this piece of metaphysical
dogma [i.e., Hume’s Dictum or MH], progress has been made. (Enoch 2011, 148)

We seem to be at an impasse. The Moorean insists that there are normative facts consistent with
AUTONOMY, INDEPENDENCE and INEFFICACY. He acknowledges that this commits him to the
existence of brute normative facts that entail necessary relations between distinct existences. But he
denies that this is too high a price to pay. And, indeed, if the only thing standing between the
Moorean and success is a piece of “dogma” backed only by vague metaphysical intuitions, it is not
hard to see why he would believe the balance of reasons favors his view.

Personally, I am inclined to think that something like MH is true. But let us just grant, for
the moment, that the price of violating MH is not so high as to rule out the Moorean view. Still,
there is more to say. As Enoch himself points out, the claim that the Moorean cannot explain
normative supervenience is ambiguous. On the one hand, it can be understood along the lines
discussed above: as the charge that Mooreans cannot explain why normative properties supervene
on the particular non-normative properties they do. It can also be understood, however, as the
charge that Mooreans cannot explain why the normative supervenes on the non-normative in general.
Enoch refers to these as “specific supervenience” and “general supervenience,” respectively. We
have already discussed Enoch’s response to the former. His response to the latter is comparatively
simple: He holds that the Moorean is welcome, just like everyone else, to appeal to the conceptual
necessity of general supervenience.

---

39 One (I think, appealing) option would be to push for an epistemological analogue—that we must have
non-circular, non-question-begging reasons for believing in relations between distinct existences, i.e. that such
relations cannot be “epistemically primitive.”
Normative supervenience is, indeed, widely-regarded as a conceptual truth. Even those who are suspicious of analyticity tend to agree that supervenience is a constraint of some important kind on our normative theorizing. But that does not mean Enoch’s response is satisfactory. Just because something is a conceptual truth does not necessarily mean that nothing informative can be said about it. And so it may be that Blackburn’s real worry is not that Mooreans cannot accept general supervenience, but rather that they cannot account for it. If they need to be able to, or if this is something they cannot do but that their opponents can, this would at least lose them some more “plausibility points.”

It is not hard to see why one would be suspicious of the Moorean’s ability to account for normative supervenience. After all, for the Moorean, normative thought and language concern an independent realm of normative facts. Out there in Plato’s Heaven, goodness sits, waiting to “attach itself” to various bits of the non-normative world. Apparently, it attaches itself “consistently,” such that non-normatively identical things will likewise be normatively identical. It’s rather hard to see why this would be the case.

Notice that non-Mooreans do not necessarily have the same problem. Consider, for example, a naturalist who claims that normative properties are natural properties. In that case, of course the normative will supervene on the natural, because goodness will supervene on that natural property it is identical with! Obviously, the Moorean can make no similar claim. For him, the supervenience relation is a relation between metaphysically distinct entities.

---

40 There is a good deal of debate over the specifics of the supervenience thesis. But nearly everyone holds at least the thesis of global supervenience—roughly, that no two worlds that are non-normatively identical can differ normatively.

41 For discussion on this point, see McPherson (forthcoming).
This shows that general supervenience involves a brute, necessary relation between the normative and the non-normative just as surely as specific supervenience does. But, of course, if the Moorean is comfortable with a brute necessary relation in the latter case, it is hard to imagine why he would be less comfortable with one here.

Enoch’s reliance on a conceptual claim, however, raises a further issue. There is, as just discussed, a metaphysical worry here about why the normative supervenes on the non-normative. But there is also a semantic worry about how it is that our language came to reflect this special feature of the normative. Here in my head are the various attitudes or other mental states that constitute my competence with normative terms. Out there in Plato’s Heaven sits goodness. How in the world did I come to recognize that this abstract entity bears this particular relation to the non-normative? For that matter, how did I come to refer to goodness at all?

3. Semantics

Mooreans owe us a story about how our term “good” came to refer to Platonic goodness, and likewise for other normative terms. The Moorean’s position here is similar to his position with respect to epistemology. In the epistemic case, the Moorean is challenged to explain the correlation between normative beliefs and normative facts. Here, the Moorean is challenged to explain “a correlation between something else down here in our world (our relevant referential practices) and some other things in Plato’s heaven (normative properties and relations)” (Enoch 2011, 183).

Because of this parallel, much of what was said in §1 is relevant here. There, I argued that there cannot be normative (ultimate) explanations of the non-normative, nor non-normative (ultimate) explanations of the normative. Because of this, most possible explanations for the relation
between referential practices and normative properties are ruled out.\(^{42}\) For example, because of INEFFECTIVE and AUTONOMY, normative properties cannot themselves explain our referential practices. Likewise, because of INDEPENDENCE and AUTONOMY, our referential practices cannot explain the normative properties.\(^{43}\) And because one correlate is normative and the other non-normative, no third factor (which would presumably be either normative or non-normative itself) can explain both.

Thus, it should be clear that, ultimately, the Moorean’s position with respect to the semantic challenge will be precisely like his position with respect to the epistemological challenge. Assuming he has to explain the correlation between our referential practices and normative properties, he can only do so by appealing to a further correlation between normative properties and non-normative properties that creates a “pre-established harmony” between the correlates.\(^{44}\) He might even appeal

\(^{42}\) This depends on the assumption that the “relevant referential practices” are themselves non-normative.

This is certainly true, at least for Mooerans, who hold that the normative and the mental are metaphysically distinct. However, it is worth noting that Ralph Wedgwood’s (2007) view attempts to capture much of what is appealing about Moorean realism while accepting that “the intentional is normative,” and thus, strictly speaking, rejecting AUTONOMY, though one might take him to be committed to a sort of “wide autonomy” for a class of properties that includes both the normative and (some of) the mental. Enoch himself draws heavily on Wedgwood in his discussion of semantics, though he admits that he cannot take Wedgwood’s view wholly on board precisely because of this issue.

\(^{43}\) Though I will leave exploration of this point for another time, it is interesting to consider what effects these limitations have on the kinds of metasemantic theories (in particular, theories of reference) available to Mooreans. For example, it seems that because of the points just made about inefficacy, the Moorean is unlikely to be able to appeal to a causal theory of reference.

\(^{44}\) This is precisely what Enoch (2011, 183) suggests.
to the same exact proposal—that because survival is good, evolutionary forces have pushed us both to have largely true normative beliefs and to successfully refer to normative properties.

4. A Final Worry

Because of AUTONOMY, the Moorean cannot offer normative (ultimate) explanations of the non-normative, nor vice versa. Mooreans must accept that certain fundamental, substantive normative facts about the relations between the normative and the non-normative are brute. As discussed, this in itself may or may not constitute an objection to the view.

Mooreans face (at least) two further challenges regarding the connection between the normative realm and what goes on in our heads. First, there is the epistemological challenge to explain the connection between normative beliefs and normative facts. Second, there is the semantic challenge to explain our normative terms’ referring to normative properties. It turns out that the only way to solve these problems is to appeal ultimately to some brute relation(s) between normative and non-normative properties. This further relation will, through a sort of “pre-established harmony,” guarantee that the right sort of connection exists between what’s going on in our heads and the normative realm.

The Moorean position is apparently coherent. But is it plausible? One potential line of attack would continue with the plausibility points game, poking holes in various bits of the Moorean story. And while I think there is a good deal to be said about why various Moorean-friendly stories are problematic, I set the task of pursuing this line of attack aside. Here, in closing, I want to say something a bit more principled about why one might (perhaps should) find the Moorean position troubling.

When judging a metanormative view, it is important to ask what we are trying to do when we engage in (meta)normative theorizing. Two possibilities come to mind: One is that we are trying
to develop a theory that can vindicate our normative thought, language, practices, etc. In other words, we take certain normative and/or metanormative practices or beliefs as given, and then explain how, if the world turns out to be a particular way, those practices or beliefs are warranted.

The other is that we are trying to develop a theory that explains why and how there are facts about what we ought to do and why various things are right, wrong, good, bad, etc., without first assuming that anything is.

This distinction is, I think, at the heart of Christine Korsgaard’s (1996) (in)famous rejection of “substantive” realism. For Korsgaard, normative theorizing is about discovering what “justifies the claims that [normativity] makes on us” (Korsgaard 1996, 10). The problem with realism, she says, is that it maintains that, ultimately, no such justification can be had.

Many realists take umbrage at Korsgaard’s characterization of their view, or believe that her “constitutivism” is similarly burdened. But what my arguments show, I think, is that Korsgaard is right at least about the Moorean. The Moorean position is, and can only ever be, a vindicating one. It offers a “just so” story about the nature of reality that, if accurate, shows us why our normative beliefs might be largely true, why our normative terms might pick out normative properties, etc. What the Moorean cannot offer us, though, is any (non-circular) reason to think that his “just so” story is true.45

45 The one way around this is for the Moorean to offer an argument for the theoretical indispensability of his view. As Enoch himself admits, it is highly unlikely that such an argument could take the form of an argument to the best explanation. However, Enoch (2011, chap. 3) argues that normative facts are deliberatively indispensable, and that this is what justifies our acceptance of them. I am not entirely convinced that this is the case. Even if it is, however, this argument succeeds only in showing that normative facts of some kind are theoretically indispensable, not that Moorean realism is true. It remains to be seen whether Moorean realism is the most plausible contender among the possible forms of realism. And it seems fairly clear, for example,
Now, as the Moorean will be quick to point out, this inability is not necessarily objectionable in and of itself. After all, it may well be that some of our beliefs are going to admit only of circular justification—are going to be “epistemically primitive.” Clearly, though, for every plausible candidate for epistemic primitivity, there are many, many more beliefs that could not be primitive—it would be objectionable if these beliefs were justified only circularly. Indeed, we have been focusing on some of these claims in this chapter, such as the claim that there is a correlation between normative beliefs and normative facts (hence the force of the epistemological challenge in the first place).

Given this, it seems that the question of whether a vindicating metanormative theory is sufficient comes down, at least in part, to the question of whether or not certain normative judgements can legitimately count as epistemically primitive, because this would explain why it is not an objection to those beliefs that we have no non-circular reasons for believing in them.

There are two ways the Moorean might go here. First, he might offer a general theory of when it is appropriate to take a belief as epistemically primitive and show that it applies to the normative. Second, he might offer a sort of “partners in crime” (or “in innocence”) argument, starting from things that are legitimately epistemically primitive—like, perhaps, modus ponens—and generalizing to other cases.

that if naturalism is at all plausible it offers a better explanation. So at least until it is shown that the Moorean’s “just so” story is better grounded than the naturalist’s “reductive” story, it seems the Moorean is at a disadvantage. (The discussion to follow offers some reasons for thinking that the onus is on the Moorean, rather than the naturalist.) Also, note that such an indispensability argument will do nothing to adjudicate between competing “just so” stories. If there is indeed more than one such story possible, there will still be no (non-circular) reason for believing that any particular story (and the resulting substantive normative facts) is true over some alternative.
I know of no one who takes the former tack. A way of taking the latter, however, is suggested by Russ Shafer-Landau’s (unpublished manuscript) arguments on genealogical challenges to normative realism. Shafer-Landau focuses on belief-forming methods, rather than beliefs themselves, but the essential points are the same. There are certain belief-forming methods, such as induction, that we have only circular reasons for taking to be reliable. We apparently have no reason to believe that induction works unless we assume (roughly) that the future will be like the past. The suggestion for the Moorean is that normativity is relevantly similar, that circularity in the vindicating story for normative judgement is unobjectionable in much the same way as circularity in the vindicating story for induction.

Of course, in order for this defense to go through, it needs to be the case that normative judgement really is relevantly similar to induction. After all, presumably I can’t just make up any autonomous domain I like and assert that because it is autonomous it legitimately admits of circular vindication.

Here, I want to point out just one potential source of concern: disagreement. There are important philosophical questions about the appropriate response to actual disagreement over beliefs—especially primitive ones. And, apparently, the existence of normative disagreement represents a marked difference from the induction case, as induction enjoys near universal support.

46 I should note that while, as discussed in note 4 in this chapter, it is not entirely clear that Shafer-Landau endorses AUTONOMY, he does suggest a version of an “autonomy thesis” in his paper: “There is no independent confirmation of the reliability of our moral faculties—their reliability can be confirmed only by a showing that they have generated moral beliefs in which we have a high degree of warranted confidence” (unpublished manuscript., 7). I remain noncommittal on whether this thesis makes Shafer-Landau’s version of realism vulnerable to all of the arguments in this chapter.

47 See, for instance, Hume’s (1748) discussion of “counter-induction.”
Further, and relatedly, there are worries about *ideal* disagreement (i.e., between fully rational agents). And, certainly, it is difficult to imagine fully rational agents opposing induction in a way that it is perhaps not as difficult to imagine them denying the reliability of normative judgement.

In his own discussion of disagreement, Enoch (2011, chap. 8) admits that the possibility of *ideal* disagreement would apparently pose a threat to Moorean realism. He denies, however, that the Moorean has reason to accept that such disagreement is possible. His suggestion, in brief, is that whether we take *actual* disagreement as evidence of rational error, on the one hand, or of anti-realism, on the other, seems to depend on whether we *start* with an assumption of realism. For example, when we find that there is disagreement between two very smart mathematicians about some mathematical truth, we tend to assume that an error has been made because we start with the assumption that there is a right answer. 48 Similarly, a Moorean should see widespread disagreement not as a threat to his view, but as evidence of error.

Remember, though, that on the Moorean view, some normative judgements are epistemically primitive—we have no non-circular reasons for accepting them. If that’s true, then it seems we have every reason to think that some normative disagreements are intractable. After all, if you start out believing that liberty is more important than happiness, but I believe the reverse, and there is no more basic normative fact that we are both appealing to, then there really is nothing we can say to one another, and, perhaps, nothing that would make my belief any more or less rational than yours, since they are both primitive! Indeed, it seems the Moorean could only deny the possibility of ideal disagreement by asserting that, as a happy accident (or perhaps as the result of yet another “pre-

---

48 For relevant discussion of whether genealogical challenges to normative realism extend to mathematical realism, see Clarke-Doane (2012).
established harmony”), we all have the same primitive normative beliefs. And so, ironically, it seems that Mooreans have more reason to believe in ideal disagreement than the opposition.

5. The Verdict

So, can Moorean realism shoulder its explanatory burden? The answer is: It depends. It depends on precisely how high the theoretical costs are of positing brute, necessary relations between the non-normative and the (metaphysically distinct) normative. It depends on whether the Moorean can construct a plausible enough “just so” story (perhaps like the evolutionary ones discussed here) to vindicate our normative thought, language and practices. And it depends, perhaps most importantly, on whether such vindication is really all we’re after. As to this final point, perhaps we should be given pause by the fact that unlike in most areas where, arguably, we take some of our beliefs to be epistemically primitive (like induction), the beliefs we are trying to vindicate seem rather frequently—and intractably—to differ.
CHAPTER THREE

HOW TO BE A NORMATIVE REALIST

Terrence Horgan and Mark Timmons (e.g., 1991 and 1992) think they have normative naturalism dead to rights. They are confident that Moore’s (1903) Open Question Argument defeats analytic naturalism. And they think that they can construct a version of their Normative Twin Earth (NTE) thought experiment to undermine any proposed version of synthetic naturalism.¹

 Unfortunately, there is little agreement about what makes something “natural.” And even if there were, it is dubious that NTE is best understood as an objection to naturalism, on any plausible definition of the natural. For example, as I discuss further on, NTE can be wielded against views on which normative properties reduce in terms of something psychological, like desires. Whether NTE undermines such a view has nothing to do with whether dualism about the mental turns out to be true. Yet if dualism is true, we might well deny that desires are “natural,” and so it is unlikely that such normative views would turn out to be forms of “naturalism.” Accordingly, one goal of this chapter is to characterize the distinctive class of views that is the appropriate target of Horgan and Timmons’ arguments.

¹ Horgan and Timmons generally speak more specifically about Moral Twin Earth. But while there are interesting issues about the moral case in particular, it seems clear that the concerns raised by Moral Twin Earth are raised in virtue of the normative nature of the moral. Merli (2002), for example, argues that it might be that the Earth word “right” is better translated as the Twin-Earth word “honorable.” Nevertheless, there is genuine disagreement between us and Twin-Earthers because we are both concerned ultimately with the question of what to do. In any case, I am concerned herein with normativity more generally, and thus refer to Normative Twin Earth.
Having proposed an appropriate target, I further consider whether it is possible to construct a view that stands in the target class, but is immune to Horgan and Timmons’ arguments. Conveniently, Horgan and Timmons (2000) offer a recipe for constructing such a view, though they express serious doubts that it can be followed to produce a successful theory. I suggest that while their recipe is largely accurate and their doubts understandable, there is room for hope. Furthermore, I suggest that if a view along these lines could be successful, it would not only avoid the problems raised by Horgan and Timmons, but would, in important ways, combine the advantages of a number of opposing views across the metanormative landscape. I close with an exploratory sketch of such a view.

1. Moore’s Case Against Naturalism

Most people agree that something is a bachelor if and only if it is an unmarried man (or something quite close to this), and that this is true just in virtue of the meanings of the words “bachelor,” “unmarried,” etc. Suppose one were to reframe this claim as a question: “Tom is a bachelor, but is he an unmarried man?” Contrast this question with another, such as: “Tom is a bachelor, but is he friendly?” The first question seems “closed.” It is not just that we all believe that Tom is unmarried; it seems that the matter is settled (a priori) just by his being a bachelor. If someone were to answer “no” to this question, we would assume that he did not really understand it. This is evidence that the relation between being a bachelor and being an unmarried man is one of conceptual necessity. The second question is not like this. You might know Tom quite well and know that he is, indeed, friendly. But the question nevertheless seems “open.” Tom’s friendliness is not settled by his being a bachelor alone; he might well have been quite unfriendly, despite his bachelorhood. It is perfectly possible that someone might answer “no” to this question, without our suspecting that he has
misunderstood it. This is evidence that the relation between being a bachelor and being friendly (if there is one) is not conceptual.

With an intuitive grasp of the open/closed distinction in hand, we can rehearse G.E. Moore’s (1903) famous Open Question Argument (OQA) against naturalism. Consider a normative property like goodness. Now consider something you think is good, or that you take goodness to supervene on—say pleasure or the property of being pleasurable. Next, ask: “I know that X is pleasurable, but is it good?” According to Moore, no matter what normative/natural property pair you select, the question will have that “open” feel. This is evidence that for all normative/natural property pairs, something’s having the natural property in question does not alone settle its having the normative property in question. And this, in turn, indicates that there is no conceptually necessary relation between any particular normative and natural properties.

Moore, of course, thought that this proved that normative properties could not be natural properties. This is partly because Moore (along with nearly everyone else at the time) accepted (or, perhaps better, assumed) a descriptive theory of reference, according to which a term’s intension (or “sense” in Fregean terms)—the implicit or explicit descriptions competent speakers associate with that term a priori—fully determines the term’s extension or reference. On this view, normative terms could have naturalistic referents only if they have naturalistic definitions. And if normative terms don’t have naturalistic referents, presumably the properties those terms name (or “express” in Fregean terms) are not natural properties.2

2 Given that we will be discussing issues in semantics, I will generally be talking about terms and their definitions, rather than concepts and their analyses. While I do not mean to commit myself to any robust view about the connection between terms and concepts, I do generally assume that, for example, if “good” can(not) be defined in naturalistic terms, then the concept ‘good’ does (not) have a naturalistic analysis.
Nowadays, the descriptive theory of reference has largely been abandoned (though it is closely related to modern two-dimensionalism\(^3\)). Indeed, modern developments in metasemantic theory have led to a number of views that develop “reductive” or “naturalistic” accounts of the normative that seem to avoid Moore’s argument entirely. Likely the most famous example, and the one originally targeted by Horgan and Timmons, is Richard Boyd’s (1988) view, which draws heavily on the analogy between the normative case and the canonical example of non-analytic (synthetic) reduction:

\[ \text{WATER. Water is H}_2\text{O.} \]

On (a very rough caricature of) Boyd’s view, normative properties are those natural properties that regulate our use of normative terms. After all, that’s (arguably) how WATER works: We learn that water reduces to H\(_2\)O when we learn that H\(_2\)O is the stuff that has and does regulate our water-talk.\(^4\)

2. Enter Twin Earth

Hilary Putnam (1975) famously argues that “meaning just ain’t in the head”—an argument that plays a key role in rejection of the descriptive theory of reference and the discovery of synthetic reduction. Putnam’s famous thought experiment, “Twin Earth,” asks us to imagine a world that is identical to Earth in all respects but one: The stuff in their lakes, oceans, etc. is not H\(_2\)O, but is rather a different substance, XYZ (that behaves very much like H\(_2\)O). He then asks: Does the Twin-English word “water” have the same meaning as our English word “water?” Most think not. After all, use of the

\(^3\) For a helpful overview, see Chalmers (2006). For application to ethics, see Jackson (2000).

\(^4\) Boyd called his view “non-reductive,” though he may have meant only that it was not analytically reductive. For the moment, I will continue to refer to these views as reductive, but (as I soon acknowledge) this is problematic.
Twin-English word “water” tracks XYZ, while use of our word tracks H$_2$O. If we were to get into an argument with our Twin-Earth doppelgangers about whether the stuff in a particular glass was water, we would presumably be talking past one another, not genuinely disagreeing.

Part of what this thought experiment shows us is that our word “water” is a rigid designator—it refers to the same thing (H$_2$O) across possible worlds.$^5$ Because of this, most people conclude that the property of being water is reducible to (which, on most views, means it is identical with$^6$) the property of being H$_2$O, even though “water” is apparently not definable in terms of H$_2$O (nor vice versa). So here we have an example where, despite there being no conceptually necessary connection between two properties, there is a metaphysically necessary connection between them—and a synthetic, rather than an analytic, reduction. Boyd’s suggestion is that there might be room for an analogous normative view—one according to which normative terms co-refer with non-normative ones, even if there is no conceptual link between the properties in question (and so nothing for the OQA to target).

Horgan and Timmons respond by using the Twin Earth thought experiment to develop a sort of metasemantic OQA. Moore’s original OQA gains force from the fact that, in general, we take ourselves to have a fairly accurate implicit understanding of the semantic value of our terms. Competent speakers will assert that questions about natural things having normative properties are open. This is evidence that all such questions really are open, and thus that normative terms do not have naturalistic definitions. But we also take ourselves to have an implicit understanding of what fixes the semantic value of our terms. Competent speakers will seemingly assert that our word

$^5$ “Rigid designator” is Kripke’s (1980) term.

$^6$ Though for an opposing (and I think, superior, though not precisely correct) view on the nature of reduction, see Schroeder (2005).
“water” and the Twin Earth word “water” have different meanings, merely in virtue of the fact that they have different causal regulators. This is evidence that causal regulation plays a role in fixing the semantic value of “water”—specifically, that it fixes its extension. This is why Twin Earth provides support for the reduction of water to H₂O.

So, say Horgan and Timmons, if normative terms’ extension is fixed in the same way that the extension of “water” is, we should expect competent speakers to have similar intuitions about a Normative Twin Earth as we do about Putnam’s. Here goes: Imagine that it turns out that here on Earth, our use of the word “good” is regulated by things that promote pleasure. Now imagine NTE, which is identical to Earth in all respects but one: Their use of the word “good” is regulated by wills that treat others as ends in themselves. If the property of goodness reduced to the property of being pleasure-promoting in the way that water reduces to H₂O, then it seems our judgements about NTE should mirror our judgements about Putnam’s Twin Earth. But they do not. For if we disagree with our NTE doppelgangers about whether something is good or not, it doesn’t seem that we are talking past one another; rather, we are having a genuine normative disagreement.

If Horgan and Timmons were merely trying to defeat the idea that normative terms refer to the properties that regulate their use, as (apparently) in the case of “water,” this would be a slam dunk. But the real question, of course, is whether they can construct a NTE counterexample for any proposed naturalist view, as they claim. What we can now see is that this is, in essence, the same as asking whether all questions of the following form are open: “I know that normative term N bears (proposed naturalist-friendly reference-fixing) relation R to stuff S, but is S the extension of N?”

7 Of course, one could have just run a straightforward OQA at the metasemantic level. For example, Putnam could ask: “I know our use of ‘water’ is causally regulated by H₂O, but is water H₂O?” and Horgan and Timmons could ask “I know our use of ‘right’ is causally regulated by pleasure-promotion, but is being right the same as promoting pleasure?” If one understood these questions well enough, one might note that the
With things framed in this way, we might just say that Horgan and Timmons’ appropriate targets are those views that propose naturalist-friendly reference-fixing relations for normative terms. Unfortunately, in order to say more precisely which views those are, it seems we would need to know which reference-fixing relations are friendly to naturalism, and thus, in turn, what naturalism itself is. But as mentioned above, not only is there very little agreement on this latter point, but it’s not at all clear that we should characterize Horgan and Timmons’ appropriate target in terms of naturalism on any plausible understanding of “the natural.” In the next section, I say a bit more about why this is the case. I then discuss the oft-pursued option of focusing on normative reduction instead of (or in addition to) naturalism. Unfortunately, I argue, taking Horgan and Timmons’ appropriate targets to be reductive requires us to make some controversial assumptions about the connections between language and reality. These assumptions may not be false, but it would behoove us to develop a more ecumenical metanormative taxonomy. In the section that follows, I do just that.

former question seems closed while the latter seems open. But this would not be a very perspicuous or rhetorically forceful way of putting the point; the Twin Earth device is helpful because rather than asking about causal regulation or reference-fixing relations, it simply asks us whether two terms have the same meaning. If the only difference between those terms is what causally regulates their use, and if competent speakers think they have different meanings, this indicates that those speakers would also be disposed to find the relevant metasemantic question closed. This, further, would be evidence that causal regulation helps determine meaning in that case. On the other hand, if competent speakers think that the words have the same meaning, this is evidence that causal regulation does not play a role in determining meaning in that case.
3. Targeting, First Pass: Naturalism and Reductive Realism

In the opening paragraphs, I suggested that “naturalism” is not the best way to characterize the appropriate target of Horgan and Timmons’ arguments. In addition to the contentious nature of “naturalism” this is, in part, because NTE seems just as forceful against certain non-naturalistic views as it does against naturalistic ones. Suppose, again, that on Boyd’s view it turns out that what causally regulates our use of normative terms is something having to do with pleasure. If pleasure turns out to be non-natural—e.g., if dualism about the mental turns out to be true—the view is no less a target for NTE.8

It is worth noting that the same goes for Moore’s original OQA. A proposed definition of “right” in terms of pleasure is no less a target for the OQA if it turns out that pleasure is non-natural. For Moore (1903, sec. 75), it is not merely that normative properties are non-natural, but that they are “autonomous”—different from properties of any other kind. And so it seems clear that whether a view is naturalistic or not does not settle the question of whether Moore means to rule it out.

One tempting alternative is to hold that both Moore’s and Horgan and Timmons’ appropriate target is normative reduction. Indeed, naturalism and reduction are frequently run together in the metanormative literature. But we have to be careful here; an important lesson to take from the last century’s discussions of the OQA is that we must not confuse (meta)semantic ...

---

8 As David Copp and Sara Worley pointed out to me independently, one might think that even if dualism were true, we would still take the mental to be natural, just not physical. Of course, this just gets us back to the problem of defining “naturalism”—more evidence that we might want to avoid using this label where possible. In any case, it seems clear that there might be understandings of “the natural” on which dualistic mental properties would be non-natural, and yet on such accounts the relevant normative views would be no more palatable to Moore or Horgan and Timmons.
arguments for metaphysical ones. To the extent that it is successful, the OQA gives us reason to think that normative terms do not have naturalistic definitions. But as we have seen, this alone is not enough to rule out the possibility that normative terms co-refer with certain non-normative ones, much as “water” and “H₂0” co-refer.

Notice that the question of whether any normative and non-normative terms co-refer is still an issue of semantics. Nevertheless, it’s not hard to see how one might be tempted to draw metaphysical conclusions from an argument for co-reference. First, it is tempting to think that an argument for such co-reference is *necessary* to argue for reduction of a normative property.⁹ For it is hard to see how one could argue for reduction of a normative property without appealing to facts about the semantics for the term(s) we use to refer to that property. Indeed, Horgan and Timmons presumably believe something like this. If not, it’s not at all clear why they would think our intuitions about NTE are so important. Such intuitions, much like those about the original OQA, would turn out to be decisive only against a particular semantics-based class of reductive views (if that).

It also seems clear that co-reference is *sufficient* for reduction—that if two terms co-refer, this means that they name the same property. And so we might think that we can generalize from the story that’s typically told to explain why WATER is true: There is a causal relation that fixes the extension of “water.” That extension is H₂0. This amounts to an identity of the “everyday substance” kind water to a molecular kind, H₂0—and thus a reduction of the former to the latter.¹⁰

However, setting aside the question of the relation between reduction and property identity, it is controversial that this argument for property identity generalizes, because it is controversial

---

⁹ Why this is tempting, and whether it is accurate, are interesting and important issues, but unfortunately not ones I can more fully explore here.

¹⁰ See, e.g., Kripke (1980). This is, of course, only a rough version of the story, but it will do for our purposes.
whether the fact that two terms pick out the same set of particulars—even across possible worlds—entails that the terms name the same property. This is in part because, on contemporary views, a term’s extension might be fixed by a relation that is independent of those aspects of the term’s semantic content that are relevant to determining what is essential to the property it names. To take a classic example: One might think that the property of being a triangle is distinct from the property of being a trilateral, because what is essential to each differs (having three angles and having three sides, respectively). Nevertheless, these properties are necessarily coinstantiated, so “triangle” and “trilateral” are coextensive across possible worlds. If we maintain that the property of being a triangle is identical to the property of being a member of the set of all (possible) triangles, and that the property of being a trilateral is identical to the property of being a member of the set of all (possible) trilaterals, then by the transitivity of identity, trilangularity and trilaterality are identical as well.\footnote{Notice that if certain aspects of geometry are analytic, this example would pose a problem for the relevant understanding of reduction even on a classical descriptive metasemantic theory.}

There are other relevant issues here, some metaphysical or metasemantic, some merely terminological. But the bottom line is that calling Horgan and Timmons’ targets “reductive” views would, at best, lead to confusion and, at worst, leave out views that are appropriate targets for their arguments. This is because, first, it is plausible but not necessarily the case that a semantic argument is required to establish a metaphysical reduction. And it is because, second, it is not entirely clear that two terms’ picking out the same set of particulars—even necessarily—guarantees that they name the same property.
4. Targeting, Second Pass: Metasemantics

At the end of §2, I argued that NTE can be seen as a kind of metasemantic version of the OQA. We ask: “I know that normative term N bears (proposed naturalist-friendly reference-fixing) relation R to stuff S, but is S the extension of N?” If this question is always open, then it seems that no R is the correct reference-fixing relation for any normative term. Given this, I suggested that we might take Horgan and Timmon’s appropriate target to be those views that maintain that this question is closed (in the affirmative) for some naturalist-friendly reference-fixing relation(s). But I also pointed out that this characterization is problematic, because it forces us to consider what makes a reference-fixing relation natural; and not only is “naturalism” a contested term but, whatever naturalism is, it isn’t really what’s at issue. One tempting alternative, as suggested by our discussion in the last section, is just to replace “naturalist-friendly” with “reduction-friendly.” But I’ve argued that reduction isn’t the right target, either.

Nevertheless, there is something right about this characterization of Horgan and Timmons’ appropriate target. Given that their original target is Boyd’s proposal of a causal theory of reference for normative terms, it makes sense that we should be looking at views that propose a reference-fixing relation of some particular kind for normative terms. What would be useful, then, is to give a characterization of this kind without referring back to metaphysical categories like “natural” or “reducible.”

To help develop such a characterization, let us return to Moore and the OQA. As discussed above, the OQA apparently rules out naturalistic definitions for normative terms. But it also might rule out certain non-naturalistic definitions, such as psychological ones (if dualism is true, perhaps). Of course, one possibility is that the OQA is supposed to rule out all definitions of normative terms. And, indeed, Moore did think that “good” was undefinable (at least on one reading). But this
does not mean that all normative terms are. For instance, Moore might actually have accepted that, by definition, something is right if and only if it promotes the good.

What makes the relevant definition of “right” immune to the OQA? The answer is not hard to find. Moore thought that normative properties were unlike properties of any other kind. But if the definiens for some normative term were describable in wholly non-normative terms, this would apparently not be the case. For if such a non-normative definition were available, there would apparently be a conceptually necessary connection between being right and satisfying the relevant non-normative description. The upshot, then, is that the OQA is probably best seen as an argument against the possibility of a wholly non-normative definition of any normative term.

NTE limits the kinds of reference-fixing relation, rather than the kinds of definition, that might be appropriate for normative terms. Otherwise, however, similar points apply: NTE rules out naturalistic reference-fixing relations. But it also rules out certain others. The ones it rules out are those on which we can offer a purely non-normative description of what fixes the extension of some normative term(s). So, I submit, what NTE seeks to rule out are reference-fixing relations that are describable in wholly non-normative terms.

---

12 I.e., the description would have to use only non-normative terms, though it might mention normative ones. For example, ‘X is good iff X is what Bob calls “good”’ would clearly be a target for the OQA.

13 So why do Moore and others bother talking about “naturalism?” This is another area where the focus on metaphysics has complicated things (unnecessarily, I think). Obviously, Moore couldn’t say that the conclusion of the OQA is that normative properties are not non-normative properties; that’s a truism!

14 It is tempting to think the issue is what the terms refer to, rather than the nature of the reference-fixing relation. To help see that it is the nature of the relation that is at issue, notice that even non-naturalists may not be able to accept a causal theory of reference for normative terms without being subject to a NTE counterexample. Suppose our term “good” is causally regulated by Platonic goodness, and NTE’s is regulated
Let us carve things up accordingly. Call a reference-fixing relation normative if normative terms are required to describe it. Call a reference-fixing relation non-normative if it can be described in wholly non-normative terms. Next, let us refer to those who are committed to a normative reference-fixing relation for normative terms *Mooreans*, and those who accept the possibility of a non-normative reference-fixing relation for (at least some) normative terms *non-Mooreans*. Finally, we add that *realists* are those in each camp that assert that normative terms in fact refer, whereas *anti-realists* are those that deny this (or deny that there is any reference-fixing relation for normative terms in the first place as, for example, certain non-cognitivists might hold).\(^{15}\) We might, of course, by something else. Do English “good” and Twin-English “good” still mean the same thing? Do we disagree with our doppelgangers about what’s good? If so, this would seem to undermine such a causal theory. And this means that it’s the nature of the reference-fixing relation, not the nature of the referent, that’s really at issue. (This example is borrowed from Mitchell (2011), who attempts to extend the NTE program as an argument against non-naturalism.)

\(^{15}\) One might argue that I’m making the same mistake I’ve attributed to others, of hastily drawing metaphysical conclusions from a semantic argument, since “realism” is typically seen as a metaphysical thesis, not a metasemantic one. Perhaps that’s right; but it seems to me that the two run together well enough, at least with respect to the case of referential failure: If you think normative terms don’t refer, it would be very surprising if you thought there were real normative facts. The case of referential success is more interesting: There are certainly those who think that normative terms refer but do not want to commit themselves metaphysically or ontologically (e.g., quietists and some constructivists). However, given that many of these people still call themselves realists, it seems we might do better considering “realism” a metasemantic category rather than a metaphysical one. In any case, I will not be talking about the realist/anti-realist divide much further, so it will not affect my argument. If others want to take my points seriously and propose that we instead refer to these metasemantically distinguished views as, say, referentialism and anti-referentialism, I have no objections.
draw further distinctions. But for our purposes, these are sufficient. I conclude that the appropriate target of NTE is non-Moorean realism (NMR).16

**5. Can Non-Moorean Realism Survive Normative Twin Earth?**

If I am correct in how I’ve carved things up, then the right question for us to ask is whether there are any referentially successful, non-normative reference-fixing relations for any normative terms that do not run afoul of our intuitions about NTE. Helpfully, in an exchange with David Copp, Horgan and Timmons lay out more precisely why they are not optimistic about the prospects for such an account.

The exchange begins with a piece from Copp (2000), arguing that while NTE is a problem for Boyd-style NMR, it is not a problem for NMR more generally.17 Boyd’s account relies on an analogy with WATER. For Boyd, the reference-fixing relation for normative terms is, at least in part, a causal one. Unfortunately, the idea that our normative terms’ extension is fixed by such a causal relation does indeed run afoul of NTE intuitions. However, Copp argues, Boyd’s (purported) heavy reliance on causation is not the best way to develop the more general metasemantic view, originally proposed by Putnam (e.g., 1975), from which Boyd draws his account. Putnam, Copp points out, acknowledged that other things in addition to the relevant causal relation, like the referential intentions and interests of those using the term in question, determine its extension. Indeed, even in the case of WATER it seems that the reference-fixing relation is not purely causal. Suppose it turned

---

16 Importantly, I think that classifying things this way helps explain why NTE can arguably be wielded against (certain versions of) expressivism and non-naturalism. See, respectively Merli (2008) and Mitchell (2011).

17 Of course, Copp (and Horgan and Timmons, in their reply) talk about “naturalism” rather than NMR; but if my arguments above have been successful then NMR is a better way to understand the class of views under consideration.
out that what causally regulates our use of the term “water” is not H$_2$O, but rather some sort of mass hallucination. Would that mass hallucination be the referent of “water?” Arguably, no. We would instead say that there is no water. This indicates that a purely causal theory of reference, according to which X is water if and only if X is that which causally regulates our use of the term “water,” is false. And it is false (in part) because it is a pre-theoretical constraint on any theory of reference for “water” that it pick out stuff that is sufficiently *watery.*

Similarly, we might suggest that there are certain constraints on a theory of reference for normative terms. We might think, for example, that when we talk about the good, we intend to refer to something that has to do with human flourishing, broadly speaking. And so if it turns out that here on Earth “good” is causally regulated by something that does have to do with such flourishing, but on NTE it does not, then this would indicate not that we and our doppelgangers are talking past one another, but perhaps instead that they are in error, because the causal regulator for *their* word “good” does not meet the standards set by our shared referential intentions.

In their response, Horgan and Timmons (2000) essentially agree with Copp about the possibility of a less causally reliant version of NMR. Indeed, they claim that they never meant to deny any such possibility. Rather, they maintain that their real concern is that for any reference-fixing relation naturalists propose, *either* it will fall prey to a NTE counterexample, as in Boyd’s case, *or* it will simply not be robust enough to get us a sufficiently determinate normative view—there will simply be too many non-normative properties that could count as the referents of “good,” “bad,” etc.

We can get a sense for why they might think this by looking at the causal theory of reference and the descriptive theory of reference as being on opposite ends of a spectrum with purely

---

18 This is reminiscent of Putnam’s (1962) famous example about Martian robot cats.
"externalist" views on one end and purely "internalist" views on the other. There are, it seems, certain constraints on a reference-fixing relation—e.g., again, that for "water," the theory pick out something sufficiently "watery"—that are transparent to competent language-users. The descriptive theory of reference takes this to an extreme, suggesting that, for each term, these constraints provide necessary and sufficient conditions that constitute that term’s intension, and that we can move directly from intension to extension. This is why descriptive theories are "internalist," because they appeal only to what’s going on in people’s heads. On the opposite end of the spectrum is (for one) a purely causal view. It takes "external" facts about how the world regulates our use of a term as dictating that term’s extension without any regard for internal features like speaker intentions.

Most actual contemporary views are somewhere in between. Arguably, the closest thing to pure internalism in the current literature is two-dimensionalism. Two-dimensionalists hold on to the idea that every term has an intension. But they also acknowledge that external facts about the actual world come into play to fill in the "gaps" in these intensions. For example, if the intension of "water" is "that liquid in the lakes, streams, etc.," we still need the actual world to tell us what that liquid actually is. And, of course, it turns out to be H₂O.

Importantly, even those who reject two-dimensionalism tend to agree that there can be a priori, pre-theoretical constraints like the "watery" one mentioned above. For example, in explicitly arguing against two-dimensionalism, Alex Byrne and James Pryor write:

[P]erhaps anyone who understands Gödel knows that it refers to a sentient being, if it refers at all. If so, the conditional If it turns out that there are no sentient beings, then it will turn out that Gödel does not exist will be a priori. Competent speakers may also know some interesting sufficient conditions for being the referent of Gödel. For example, if competent speakers know the necessary condition just mentioned, then they will also know that if there is exactly one sentient being and if
Gödel refers, then it refers to this sentient being. If so, the conditional *If it turns out that Gödel exists and there is exactly one sentient being, then it will turn out that Gödel is this sentient being* will be a priori. (Byrne and Pryor 2006, 50)

Given this, it seems unlikely that any view will be purely externalist. The right reference-fixing theory for a given term will include some “internal” factors, including relevant a priori constraints, as well as, at least in some cases, “external” favors such as what causally regulates use of that term.  

We are now in a position to see how Copp’s suggestion is supposed to work. Suppose our term “good” tracks property A and NTE “good” tracks property B. On the causal theory of reference, our word refers to A and their word refers to B. This means either that our words have different meanings or that they have the same meaning but different referents. But our semantic intuitions seem to tell us that they have the same meaning. And as both Copp and Horgan and

---

19 This point is made stronger by the fact that if anything were to have a purely “external” reference-fixing relation, it would likely be proper names, which seem to be introduced solely for the purpose of referring to some particular individual. If even *these* have certain minimal a priori constraints, it seems all the more likely that all terms will. I do not rely on this point in what follows, however—only on the assumption that *some* terms have such constraints, and that normative terms are among them.

20 And “good” also plays the same role with respect to motivation, assessment, etc. on NTE that it does here. Without this caveat, the intuition that Earth “good” and NTE “good” have the same meaning becomes suspect. What if, for example, those on NTE use “good” to refer to pleasure but aren’t thereby motivated to seek pleasure? In that case, we might deny that their word was the same as ours, *even if* pleasure was the causal regulator for *both* our terms!

21 Obviously, whether this is possible or not depends on how we understand what meaning is.
Timmons discuss, the latter option is unappealing because it involves an apparently objectionable form of relativism.\textsuperscript{22}

But, Copp points out, there might be a priori constraints on the correct theory of reference. Those constraints might rule out A (or B) as the referent of “good.” In that case, we wouldn’t end either with a difference in meaning or objectionable relativism. Rather, we would conclude that our terms mean the same thing but that those who have been using the term to refer to B (or A) are simply in error.

Once we understand the suggestion, it is not hard to see why Horgan and Timmons would be dubious. Any proposed a priori constraint will be either normative or non-normative. Normative constraints (as, they argue, the stuff about “flourishing” probably would be) don’t help, because we then need a further reference-fixing story for that term. Non-normative constraints, on the other hand, lead to referential indeterminacy:

Given that the groups [on Earth and NTE] are similar in the ways we stipulate in our argument, facts about these groups together with the sorts of referential intentions and interests that Copp describes do not suffice to pin down the referents of moral terms – so we have maintained and Copp has done nothing to challenge this claim. Both on earth and on Twin Earth, too many natural properties are available as eligible referents for moral terms – for instance, both consequentialist-functional properties and deontological-functional properties. Out of the frying pan of objectionable

\textsuperscript{22} Technically, such relativism would count as a successful version of NMR, in that normative terms would refer and the reference-fixing relation would be non-normative. Aside from the plethora of other objections to relativism, it seems clear that this kind of referential success is too thin to make good on the kind of “objectivity” intuitions that ground most people’s realism.
relativism; into the fire of referential indeterminacy! (Horgan and Timmons 2000, 149)

Why should we think that these a priori constraints will not “suffice to pin down the referents of moral terms?” The answer can be found by looking back to the original OQA. If Moore is right, then there are no non-normative definitions for normative terms. But if what we know about (say) “good” a priori isn’t enough to give us, in non-normative terms, necessary and sufficient conditions for being good, it is hard to see how what we know a priori could tell us what the extension of “good” is. If that’s right, then any plausible reference-fixing relation will have to appeal to external factors like causal regulation. And Horgan and Timmons are confident that as soon as such external factors come in, NTE will show that the proposal fails.23

I am inclined to agree with Horgan and Timmons that any such external appeal will fail. If that’s right, then the only way to defend NMR would be to tell a reference-fixing story that appeals only to a priori constraints—i.e., that is purely internal. And, again, given the force of the OQA, it is tempting to think that this cannot be done.

23 It should be noted that while some people accept that there is some indeterminacy in the extension of normative terms, this is not the kind of indeterminacy Horgan and Timmons are worried about. For example, Shafer-Landau (1994) points out that there might be normative indeterminacy that’s analogous to what we find in Sorites paradox cases. There might be some “gray area” indeterminacy in the extension of “good” just as there is in the extension of “bald” or “heap.” Nothing Horgan and Timmons say about Copp’s argument commits them to rejecting this kind of indeterminacy. Their concern, rather is with a kind of “extensional inadequacy” where the extension of normative terms is not determinate enough to meet some minimal threshold for developing a satisfactory normative theory. Thanks to David Shoemaker for pointing out the relevance of this distinction.
It is important to see, however, that the OQA alone is not sufficient to rule out the possibility of an a priori, non-normative reference-fixing relation for normative terms. Recall the Gödel case above. The point there was that while there may well be no a priori definitions for names, there might nevertheless be circumstances in which what we do know about a name a priori (such as that it necessarily refers to a sentient entity if it refers at all) is sufficient to fix its extension. If something analogous is true in the normative case, we may be able to fix normative terms’ extension a priori without the need for an a priori definition.

The obvious worry here is that the Gödel case smuggles in a posteriori information about the world in question. After all, we only know that “Gödel” refers to that one sentient being if we already know that there is only one sentient being and that “Gödel” refers—neither of which is a priori. If normative cases are like this, and we need certain information about the world in order to fix the referents of our normative terms, it seems Horgan and Timmons can once again construct a relevant NTE. Luckily, however, the normative case is not like this. To see why, let us turn to a recent argument from Christian Coons (2011).

6. Coons’ Supervenience Argument

I have suggested that there are a priori constraints for many, if not all, terms. It is tempting to call these conceptual truths, though I will avoid this phrasing for obvious dialectical reasons. What is important is that nearly everyone can accept that there are such constraints. A committed Quinean might balk at the absolutist implications of the word “constraint.” But even he should acknowledge that for many terms, there are certain claims that are very near the center of our “web of beliefs”.

---

24 This is an allusion, of course, to Quine (1951).
that are nearly indistinguishable from full-blown constraints for all practical purposes. In Coons’ argument, he relies on perhaps the most widely-accepted candidate for such an a priori constraint: SUPERVENIENCE. No two worlds that are non-normatively identical (including with respect to their histories) can differ normatively.

Coons argues that SUPERVENIENCE has a surprising implication: In normative theorizing, we can move from certain claims about what is possibly good to claims about what is actually good. He opens with a thought experiment:

You drive down Main Street near the town auditorium. Inside, the local charity network will hold a seminar on how to contribute to local and global causes. Everyone is on their way there; dozens line the street heading towards the meeting. An approaching traffic light turns red; and the crowd streams across the crosswalk ahead. If you punch the gas, you would take out 10, seriously injure 25, and topple that 1000-year-old Spruce you swore to protect. You abhor the prospects of harming others and of repairing your car. Punching the gas also furthers no one’s interests or well-being; but is it wrong? (Coons 2011, 83–84)

Coons points out the apparently obvious fact that on any plausible normative theory, punching the gas would be wrong. Nevertheless, he says, one might still wonder whether it actually is wrong. Specifically, one might be tempted by a kind of nihilistic doubt into supposing that perhaps nothing is right or wrong. Coons argues that given the truth of SUPERVENIENCE, it seems that if we can do conditional normative theorizing—showing, for example, that if anything is wrong, then punching the gas is—then we can actually prove something substantively normative—we can further demonstrate that punching the gas is actually wrong.

The argument, briefly, is as follows: First, assume SUPERVENIENCE: no two worlds that are non-normatively identical can differ normatively. Second, assume that the instantiation of normative
properties is metaphysically possible in a world like ours. That is, there could be a world non-
-normatively identical to ours in which certain things are right or wrong, good or bad. Third, do a bit
of conditional normative theorizing—if anything is bad, punching the gas is.

Call the counterpart world that is non-normatively like ours, in which certain things are right
and wrong, W. If certain things are right or wrong in W, and punching the gas is wrong if anything
is, then punching the gas is wrong in W. SUPERVERIENCE entails that if W and the actual world are
non-normatively identical (which, by stipulation, they are), then they are also normatively identical.
So if punching the gas is wrong in W, it is wrong in the actual world. So punching the gas is wrong
in the actual world.

Now, of course, some of these premises need defense. The most vulnerable premise, it
seems, is the one which claims that there is a metaphysically possible world like ours in which certain
things are right, wrong, good or bad. But as Coons points out, while this claim is certainly not
obvious, there aren't really any good arguments against it. Indeed, even most error theorists seem to
accept the truth of this claim, given that they offer non-modal arguments, such as arguments to the
best explanation, for their conclusions. Thus, at the very least, it seems we are permitted to assume,
unless we are shown otherwise, that such a world is possible.

Second, we have to accept that we really can do such conditional normative theorizing—that
we have some insight into what would be wrong, if anything is, and that, in particular, if anything is
wrong, punching the gas is. And notice that this claim is much more plausible when it is specifically
about morality, than when it is about normativity more generally. Consider, for example, the

25 The *locus classicus*, of course, is Mackie (1977). For more recent work on error theory see, for one, Olson
(2010). Coons also discusses Richard Joyce (2002), who does argue that the instantiation of *moral* properties is
metaphysically impossible. However, Joyce’s argument relies on further normative claims, and thus cannot be
extended as an argument for *normative* error theory (as Mackie’s and Olson’s arguably can).
suggestion above that our referential intentions with respect to moral terms ensure that they will capture something about human flourishing. It seems clear that no matter how one understands flourishing, there is no way that punching the gas in the above example would contribute to such flourishing. So it might indeed seem clear that if anything is morally wrong, this is. But it is less clear that we have such constraints when we think about normativity more generally. Could it really not be that we have reason of some kind to punch the gas? My intuitions certainly balk at the suggestion that there is such a reason; but these seem like substantive normative or moral intuitions, and it is not clear that these are the right sorts of thing to constrain the reference of moral terms—i.e., that these are genuine a priori constraints like SUPERVENIENCE.  

Whether or not this particular example succeeds, it seems that Coons may have provided us with a strategy for fixing the extension of normative terms. Importantly, this strategy relies only on premises that can be described in non-normative terms, and so the precise reference-fixing relation (which I will not attempt to work out here) will be non-Moorean. The key claim—SUPERVENIENCE—is a priori and thus “internal” and so, as discussed in the last section, is unlikely to be subject to an NTE counterexample. But SUPERVENIENCE is also not a non-normative definition of any normative term, and thus does not run afoul of the OQA.

26 Though see Foot (1978). Incidentally, I think that my points here reflect a large part of what’s wrong with Jackson’s (2000) view. He proposes that there are certain a priori “subject-determining platitudes” that constitute the intensions for normative terms. The idea of a “subject-determining platitude” is very much like my proposal that there are a priori constraints on the referents of normative terms. But what Jackson is talking about, it turns out, are certain apparently substantive moral judgements about which particular things are right and wrong, rather than arguably pre-theoretical constraints like SUPERVENIENCE. Because of this his view is subject to NTE counterexamples (or ends in objectionable relativism). For discussion, see Horgan and Timmons (2009).
Unfortunately, just as Horgan and Timmons would predict, we do run up against concerns about indeterminacy. After all, even if it is true that it would be wrong to punch the gas, and to do any number of other things that all plausible normative theories agree on, it’s not at all clear that there is enough agreement between all normative theories to develop an adequate normative theory.

Coons has a reply to this (though he is not considering it in terms of Horgan and Timmons’ argument). He claims that if we get to the end of normative theorizing and find that we have a disjunction of plausible views that offer different normative advice, we should simply conclude that we have disjunctive obligations:

Perhaps even after fully idealized normative inquiry has run its course, two or more theories will remain. But that outcome would not re-introduce the threat of error theory or skepticism. On the contrary, I submit that we would have everything required for a compelling and conclusive normative theory. Suppose that only Theory A and Theory B survive idealized normative inquiry. When the verdicts of the two theories overlap, you can be confident that you must abide by their shared verdict. But what about cases where the verdicts of A and B conflict? For instance, if A tells you to do x, and B tells you to do y in some circumstance. In that case, I offer that you must conclude that you should do x or y (and not z or v, or w, [sic.])—not everything would be permitted. (Coons 2011, 93)

This may indeed be the right response where error theory is concerned. But it may not be as dialectically forceful against competing metanormative theories. I have proposed that we can use Coons’ argument to help develop a non-Moorean view where the extension of our normative terms is fixed in accordance with certain a priori claims like SUPERVENIENCE. But if the resulting view is radically impoverished, this might indicate that we should jettison the kind of non-Moorean story that is being told here for some alternative. An expressivist or a non-naturalist, for example, might
insist that because this view leaves us with too little by way of normative guidance, it would behoove us to further examine the prospects for their preferred view. And this is indeed what I would expect Horgan and Timmons to say, that if the best that NMR can offer is a proof for the genuine rightness or wrongness of a small, select set of acts that everyone (with the exception of error theorists) would agree about anyway, this is reason to try to discover other ways of getting some alternative, and more robust, normative view off the ground.

Nevertheless, while the resulting view might not be determinate enough to get us everything we want from a normative theory, it would not be wholly indeterminate. If Coons is right that the above argument against punching the gas succeeds, then this is indeed a proof of a substantive normative fact, one that proceeds from claims about the nature of normativity that are given in wholly non-normative terms (e.g., SUPERVENIENCE). We have demonstrated, given only some a priori claims about the nature of normativity (including about what would be wrong if anything is), plus certain non-normative facts about the world, such as about what punching the gas would do in the case described, that something actually is wrong. This is an important result, because if it can be done in this limited way, this certainly undermines Horgan and Timmons’ suspicion that it cannot be done at all, and paves the way for more robust discoveries along non-Moorean lines.

7. Why Non-Moorean Realism Is Attractive

In the previous section, I suggested that Coons’ argument from SUPERVENIENCE can potentially be used to develop a realist view that offers a non-normative reference-fixing relation for normative terms without falling prey to our intuitions about NTE. Worries about extensional adequacy remain,

---

27 Of course, SUPERVENIENCE includes the term “normative,” but it should be clear that this is not a normative term in the relevant sense.
but may be surmountable. In the final section of this chapter, I offer a rough and exploratory sketch of how a more robustly determinate version of NMR might be developed. First, though, I discuss why this view should be particularly attractive to those of us interested in developing a view on which there is objective normative truth. In what follows, I use “CNMR” to refer to the combination of the non-Moorean thesis that the reference-fixing relation for (some) normative terms is non-normative in combination with Coons’ argument from SUPERVENIENCE.

The attraction for those with naturalist inclinations should be obvious. Although “naturalism” is a contested term, so far as I can see there is no reason to think that CNMR is in tension with actual naturalistic commitments. After all, nothing in my arguments appeals to anything supernatural, mystical or Platonic.

I suspect that few people come to the table committed to normative reduction in the way that many are committed to normative naturalism. But those that do should also find CNMR attractive. After all, with a non-normative reference-fixing relation in hand, one can always go the “Australian” route and consider all such relations to entail reduction (and indeed, it is precisely these theorists who would be most likely to come to the table with reductive commitments in the first place).

More interesting are the ways in which CNMR should appeal to naturalism’s traditional opposition. Those on the other side of the realist camp—those I have been calling Mooreans, but who are more typically known as non-naturalists—tend to think that only their view can capture what is distinctive about normativity. David Enoch (2011), for example, argues that his version of the Moorean view is the only view that fully “takes morality [and normativity] seriously.” He has two basic reasons for thinking this. The first is that (he believes) only cognitivist views really make good on our sense that normative truth is objective. The second is that (he believes) normative properties
are just too different from everything else in the world for there to be a successful naturalistic reduction.

Like Enoch, I am a cognitivist and am inclined to think that non-cognitivism is a non-starter precisely because it cannot capture an important kind of normative objectivity. Nevertheless, it seems to me that CNMR captures some of what is appealing about the non-cognitivist project, in that its theoretical building blocks are judgements about normativity—specifically, competent language-users’ beliefs about constraints on their use of normative terms (e.g., SUPERVENIENCE). But unlike with expressivism, because the judgements in question are beliefs, CNMR avoids entirely the need to develop a totally new semantics for declarative statements that express non-cognitive attitudes (and thereby also avoids the embedding problem that is, arguably, expressivism’s greatest challenge).

I also think that in an important sense, CNMR can make good on the just-too-different intuition. This intuition is, I take it, roughly the sense that facts about what ought to be the case couldn’t just turn out to be facts about what is the case. Now, certainly, a successful version of CNMR would show that normative terms refer to a distinctive set of things that can also be described in non-normative terms. But though I said before that CNMR can thus be considered reductive, I do not think that it should. If there is any case in which we would fail to capture what is essential to a property by identifying it with the set of things it is instantiated in, it is here. What makes normative properties the properties they are arguably has something to do with things like their relevance to action, the very thing that, in the next section, I appeal to in my sketch of a more fully developed version of CNMR. What’s more, nothing said thus far does anything to undermine a non-reductive interpretation of CNMR. Coons’ argument, for example, relies explicitly on the fact

---

28 As I argued in Chapter One.
that normative properties *supervene* on non-normative ones. Nothing is said to indicate that this
relation should further be considered one of identity.

Furthermore, given that CNMR derives substantive normative facts from claims (like
*supervenience* and one about action-guidance discussed in the next section) that arguably describe
what’s distinctive about normativity, it seems CNMR makes good on our sense that normative
properties are unlike properties of other kinds. But CNMR also lacks the intensely worrying
semantic, epistemic and metaphysical commitments that lead so many to reject Moorean views.²⁹

Of course, none of this means much if we cannot make good on the CNMR program—if
we cannot specify a non-normative reference-fixing relation for at least one normative term that, in
conjunction with Coons’ argument, allows us to develop an extensionally adequate normative
theory. Thus, in closing, I offer a rough sketch of how one might develop a more robustly
determinate version of CNMR. If it—or another view along the same lines—can succeed, I think we
would end with the best of all possible metanormative worlds: We would have a genuinely realist
metanormative view that addresses the nihilist, makes good on our sense that the normative is
distinctive and avoids the traditional semantic objections to naturalism, all without appeal to anything
worryingly queer.

²⁹ As discussed in Chapter 2. I should note that therein, I characterize Moorean realism as a metaphysical
view—in terms of the “autonomy” of the normative—instead of as a semantic one in terms of the nature of
reference-fixing relations for normative terms. Dialectically speaking, this is not particularly problematic,
given that with only one possible exception I know of—Wedgwood (2007)—no one has ever pulled these
apart.
8. “The Normative Cogito”

Constitutivists believe that the facts about what we ought to do are the demands of agency. Once we understand what the nature of agency itself requires of us, we understand what the normative facts are. Now, I am not a constitutivist, and I am dubious about the prospects for developing a sufficiently robust normative view out of the demands of agency. But there are some very smart people who think we can. And if they are right, then arguably their view faces only one major challenge: They must vindicate their bedrock claim that the normative facts just are the demands of agency. This is the essence of David Enoch’s (2006 and 2010) “shmagency” objection to constitutivism: The constitutivist is committed to the claim that we ought to do as agency demands. But this is a normative claim. So it seems the constitutivist has no non-circular way of justifying her view.

Notice that constitutivism is a form of NMR, because constitutivists are committed to the existence of non-normative reference-fixing relations for some normative terms. This should be obvious given that their bedrock claim is in essence the proposal of a non-normative reference-fixing relation for a normative term, something like: X is in the extension of “normative obligation” if and only if X is a demand of agency. It should not be surprising, then, that Enoch’s objection can be recast as a sort of OQA: “X is a demand of agency, but is X something I ought to do?” seems to be an open question.

---

30 Perhaps most notably, Christine Korsgaard (e.g., 1996) and J. David Velleman (e.g., 1989).

31 Assuming “agency” and “demand” can be defined in non-normative terms. And, of course, constitutivists must think that these terms can be so defined (assuming they can be defined at all) if they are to avoid begging the question.
What I want to suggest is that we might be able to use Coons’ SUPERVENIENCE argument, along with another plausible constraint on the extension of normative terms, to vindicate the claim that we have reason to be agents. If that’s right, and if constitutivists are also right that they can develop fully determinate normative views from the demands of agency, then my doing this work provides the seed necessary to grow a full-fledged normative theory.

Now, there are many ways of understanding the nature of agency. One way—is inspired by Korsgaard (1996)—is to understand agency in terms of confronting what she calls “the normative question”—the question of what to do. To confront the normative question, it seems, means taking this question seriously—attempting, in deciding how to act, to figure out how one in fact ought to act. Thus, if it can be shown that we have reason, in deliberating, to try to figure out what we ought to do, this means that we have reason to confront the normative question, and thus to be agents.

Earlier, I suggested that SUPERVENIENCE is a plausible constraint on the extension of normative terms. Here is another:

ACTION-GUIDANCE. The normative facts must be capable of guiding the actions of (at least some) agents.

Admittedly, this is significantly more controversial than SUPERVENIENCE. But it is, I think, sufficiently plausible for my exploratory purposes in this section.

---

32 For some discussion of how to understand Korsgaard’s view on agency, see McDonald (2010). Korsgaard (2009, 32) asserts that one is only an agent to the extent that one is governed by the laws of practical reasoning, which apparently means, for one, acting as one judges one ought to act, and thus, arguably, trying to figure out how one ought to act—i.e., trying to answer the normative question.

33 Arguably, this (or something quite close to it) undergirds the more widely accepted “ought implies can.” It cannot be that one ought do something one cannot do because that fact would not be one that could guide
Recall that Coons’ SUPERVENIENCE argument allows us to move from possibility to actuality. If our conditional normative theorizing tells us that certain things are right, wrong, good, bad if anything is, then it turns out that these things are, in fact, right, wrong, etc. Thus, if we can show that:

(A) If we ought to do anything, then we ought to try to figure out what we ought to do.

It will follow from Coons’ argument that:

(B) We ought to try to figure out what we ought to do.

The question, then, is whether all plausible normative theories would hold that we ought to try to figure out what we ought to do. Now, certainly, it might sometimes be the case that we ought not try to figure out what we ought to do. It might be, for instance, that in certain circumstances deliberating at all is a bad idea. Perhaps there is an emergency that can only be dealt with if we act immediately, without considering the normative question. I do not think, however, that it could be a basic normative fact that:

one’s actions. Action-guidance might also play a role in motivating certain kinds of reasons internalism. Some, like Williams (e.g., 1981), argue (roughly) that something could be a reason for one to Φ only if it could motivate one to Φ. I am not an internalist (in this sense), but it seems that one might get to this view as a result of thinking that being motivational in this way is the only way a purported reason could guide one’s actions.

It is an interesting question how to deal with this fact. One plausible suggestion is that when we learn about the possibility of such situations, we have reason to inculcate in ourselves certain dispositions to act in particular ways under those circumstances. For example, I might have reason to develop a disposition to jump in front of a speeding car if it is bearing down on my child.

That is, it might be that given certain facts about how the world is, the reason we have to seek the normative truth is overridden—e.g., if a demon threatens all life on Earth unless I stop trying to figure out
IGNORANCE. We ought not try to figure out what we ought to do.

In order for the argument to come to work, I need two further assumptions, both of which, again, are plausible but not uncontroversial: (1) In order for a normative fact to guide our actions, we must come to know it; (2) We don’t start out knowing the normative facts, but have to learn them by trying to discover what they are.

Now, according to ACTION-GUIDANCE, the normative facts must be capable of guiding our actions. IGNORANCE tells us not to try to figure out what we ought to do. On the assumptions just made, telling us not to try to figure out what we ought to do effectively tells us not to come to know what we ought to do, which in turn means we can’t be guided by any of those facts about what we ought to do. So IGNORANCE in effect tells us not to be guided by the normative facts. But that means that IGNORANCE effectively tells us not to be guided by IGNORANCE, since it is itself a normative fact.

Now, to be fair, ACTION-GUIDANCE doesn’t tell us to be guided by the normative facts, but only that the normative facts have to be able to guide us. Still, it does not seem very plausible that normative facts must, by definition, be able to guide us, and yet that they should not. (Indeed, one might think that, plausibly, it is another a priori constraint that we ought to be guided by the normative facts.)

So, I submit that IGNORANCE cannot be a basic normative fact. I know of no direct arguments, however, that the opposite claim—that we ought (or at least, to take care of emergency cases, have reason) to try figure out what we ought to do—could not be true. And it certainly seems plausible that it could be true. So there seems to be every reason to believe that there is a possible

what I ought to do. But I do not think it is plausible that a fundamental normative truth is that we shouldn’t seek out the normative truth. Thanks to David Copp for helpful discussion on this point.
world like ours in which we have reason to try to figure out what we ought to do. And thus, further, by Coons’ SUPERVENIENCE argument, it is a normative fact in the actual world that we have reason to try to figure out what we ought to do.

Of course, this alone does not get us very far at all. But if we can indeed characterize agency in terms of taking the normative question seriously, and if taking the normative question seriously involves trying to figure out what one ought to do, then it follows that we ought to be agents. And while this alone also does not get us very far, there is a whole constitutivist tradition that claims that once you have this, you can generate a full-blown normative theory.36

9. Conclusion

In this chapter, I have attempted to clarify how the Horgan and Timmons’ anti-“naturalist” program works: by suggesting that NTE undermines all externalist, non-normative reference-fixing relations, and that all others suffer from extensional inadequacy. Whether or not the kind of NMR I have been advocating can avoid the latter concern, I submit that this is the only way forward for those hoping to offer a non-normative reference-fixing relation for normative terms. If my proposal above, or something like it, doesn’t work, Horgan and Timmons do indeed appear to have NMR over a barrel.

However, it is important to be clear about what they would have succeeded in showing, which is, again, that there can be no non-normative reference-fixing relation for normative terms. Whether this would spell doom for naturalism or reductive realism depends on whether or not these

36 Of course, constitutivists have to accept a definition of agency from which to develop their fuller view, and it remains to be seen whether, even if they can develop one from some understanding of agency, the “confronted by the normative question” definition will be sufficient.
metaphysical positions are ultimately dependent on certain (meta)semantic claims. Either way, it is important to recognize the distinction.

Hopefully, whether the version of NMR I have sketched is the right way or not, the path ahead is clearer.
BIBLIOGRAPHY


———. 1748. *An Enquiry Concerning Human Understanding.*


Moore, G.E. 1903. Principia Ethica. At the University Press.


———. unpublished manuscript. “Skorupski on Being For”. University of Southern California.

———. unpublished manuscript. “Genealogical Critiques of Normative Realism.” University of Wisconsin, Madison.


