The Unnatural Reason:
How to Open Up a Normative Ontological Gap

Some ethical statements strike me as clearly true. Torture is wrong. Kindness is a virtue. Pleasure is good. One straightforward way for an ethical statement to be true is if there are ethical facts of the matter that make it true. And if that is how we choose to spell out the truth conditions in ethics, then we will need to know what an ethical fact would be and whether there are any.

According to naturalism, our world is fully composed by the kinds of entities, properties and relations studied by the natural sciences. The natural facts fix all the facts. To keep both convictions—the conviction that some ethical statements are made true by ethical facts, and the conviction that the world is fully composed by a naturalized ontology—one must hold that the ethical facts (i.e., entities, properties or relations) are to be identified with (in some suitable way) natural entities, properties or relations. That is the way of ethical naturalism.

Here I supply an argument against normative naturalism quite generally, not just ethical naturalism. I argue that if there are suitably mind-independent normative facts of any kind there is a distinctive relational property that goes missing from the natural world, viz., the favoring relation found in normative reasons. No kind of reason fact, whether it be hypothetical and

---

1 Here I will not facts from other metaphysical players like entities, properties and relations. I will use these terms interchangeably to refer to the furniture of the world. Nothing here turns on finer distinctions.
grounded in desires, categorical and grounded in morality, or epistemic and grounded in true belief, is metaphysically reducible to the natural world.

To make the case I will first consider various ethical naturalization projects, where naturalization strategies have been most fully developed. The strategies typically develop sophisticated semantic theses, usually with regard to terms like ‘good’ and ‘right,’ to show that ethical properties, like goodness and rightness, can be identified with (in some suitable way) natural properties. I will then try to demonstrate that no such strategy will work for any kind of genuinely normative fact because normative reasons--what I take to be a basic normative unit indispensable to any normative discipline--cannot be identified with or realized by anything in the natural-phenomenal world (and a fortiori cannot be reduced to anything in the natural world). As I will understand the desired ontological reduction, a domain can be reduced if and only if the truths of that domain are a priori entailed by the complete set of the reducing truths (and nothing more). Because the complete set of natural and phenomenal truths (and nothing more), fails to a priori entail any truths about reasons, I conclude that any true reason claim requires ontology above and beyond the natural-phenomenal to make it true. Dualists in mind who rely on similar

---

2 The only issue that concerns me is ontological. The naturalization I am concerned with does not require that the methods or epistemologies of different disciplines be the same as the methods or epistemologies of the natural sciences. Nor does naturalization require that explanatory schemas concerning higher-level phenomena be eliminated in favor of the concepts and explanatory schemas of basic physics or other sciences, but only that all the ontological stuff in the world be natural stuff, perhaps thought of and talked about differently by physicist, psychologists, historians, and others, but natural stuff all the same. In essence, ethical naturalists want to prove up the suspicion that if we were to catalogue all the natural entities, properties, and relations of the world there would be no kind of ethical entity, property or relation left unaccounted for.

3 Normative reasons, as opposed to motivating or explanatory reasons. Throughout I am only concerned with normative reasons, i.e., considerations that count in favor of various actions or attitudes, to be explained below.
tests for ontological reduction, and who are realists about normativity, should embrace a third irreducible property in the world, viz., the normative relational property found in reasons.

I. Naturalization Projects

In the modern period Moore (1903) stimulated the naturalization debate when he argued that the property of intrinsic goodness is not identical to natural facts, or to any other kind of fact for that matter, and it is therefore a simple, *sui generis* property, detectable by a special faculty of moral intuition. To support the position he employed the (in)famous Open Question Argument, which works along the following lines with respect to intrinsic goodness. Take any putative definition of ‘good’ that is non-circular. For example: good is that which is desired. If the definition is correct, then Moore thought that any question of whether ‘good’ really has the same meaning as ‘that which is desired’ should be closed and should have a closed feel. But it does not seem like the question “is what is desired good?” is closed. To the contrary, it is an intelligible, open question whether what is desired really is good. (Contrast the question “is what is good good?” That certainly is a closed question.) Because Moore thought that all non-circular definitions of ‘good’ would be open he rejected any such definition, and he concluded that the property picked out by the term ‘good’ could not be the property picked out by any other term. That is, it is a simple, *sui generis* property.

This final step in the argument betrays it. The OQA most clearly shows that the meaning of the term ‘good’ is not *synonymous* with the meanings of any other terms, at least not in a way that is readily accessible to anyone who grasps the meanings involved. But it is unclear why synonymy should be a decisive test for determining whether two terms pick out the same property in the world. Even if we expand the test to a broader form of analyticity between normative and non-normative terms, the test is still too demanding. The synonymy criteria, or an
analyticity criteria, is simply not necessary to determine whether goodness is a natural property, or whether it is a simple, *sui generis* property, as Moore thought it to be.

### A. Naturalization After Moore

Recent naturalization projects try to exploit more sophisticated semantic theories to show that ethical terms can and do pick out the same properties as do natural terms despite the failure of synonymy. I would like to briefly sketch two kinds of semantic theses that allow non-synonymous terms to nevertheless refer to one and the same (natural) property in the world. The first theory is causal regulation semantics (CRS), employed by Richard Boyd (1988), and often considered to be a characteristic of Cornell realism. Under CRS, a term’s extension, and so part of its meaning, is determined by which entities, properties or relations are responsible for causally regulating our use of the term. Consider the non-ethical term ‘water.’ Under CRS, ‘water’ refers to the chemical substance H$_2$O because H$_2$O causally regulates our use of the term. A historically oriented CRS would say that sometime in our past there was a baptism of the world ‘water’ to refer to the clear, potable stuff that fills our lakes and rivers and falls from the sky during rain storms, but particular versions of CRS is not our present concern. What is important as far as ontology is concerned is that the terms ‘water’ and ‘H$_2$O’ need not enjoy any synonymy in order to refer to the same in-the-world property, and so synonymy is a poor litmus test for ontological identity.

Under CRS, ethical terms like ‘good’ and ‘right’ will also refer to features of the natural world—and will therefore require no ontology in excess of the natural one—if some natural entity, property, or relation enjoys the right causal nexus with our use of the terms.

---

4 I find other Cornell realists like Nicholas Sturgeon (1985) and David Brink (1989) less explicit about their semantics. Though some characterize all these folks as causal regulationists, Sturgeon and Brink might best be characterized as role semanticists, described below.

5 See, e.g., Kripke (1972), and Putnam (1975) for initial statements of the view.
Various versions of what I will call ‘role semantics’ also promise to naturalize ethics. Consider functional role semantics, which is represented by analytical functionalism about the lay concepts of psychology.\(^6\) There the thought is that folk terms like ‘desire,’ ‘belief,’ and ‘pain’ refer to a higher-order dispositional-functional role that these psychological states play in the lives of agents. These functional roles can be realized by any number of lower-order states that causally interact with one another, such as the neuro-chemical states found in humans, or hydraulic states found in Martians,\(^7\) or ectoplasmic states found in ghosts. If the functional roles of desire and belief are realized by natural properties in our world, then they require no additional ontology above and beyond the natural. It is no concern if these functional properties can also be realized by non-physical stuff in other worlds (e.g., ectoplasmic worlds) if we restrict ourselves to an appropriately weak version of naturalism that requires that all properties in this world be realized naturally.

In the case of normative terms like ‘good’ and ‘right,’ it is unlikely that they refer to higher order functional roles, but they might find a different kind of role in a system of platitudinous normative statements. Frank Jackson (1998, 1995 with Pettit), Philip Pettit (1995) and Michael Smith (1994) each have versions of what we can call ‘platitude role semantics,’ whereby an ethical term’s role is determined by its place in a system of platitudinous statements that make use of ‘good,’ ‘right,’ the virtue terms, and other normative terms. This system of platitudes is analogous to the interconnected functional roles that appear in folk psychology, and the

\(^6\) See, e.g., Lewis (1972), Armstrong (1968).
\(^7\) Lewis (1980) defends a kind of type-type identity theory, according to which human pain is a different type of property than Martian pain. Each of these properties is then type identical to certain brain states (in the case of human pain) or hydraulic states (in the case of Martians). I think type-type identities will only get us so far, and the better view is token identity theory, according to which pain tout court is not type identical to any physical property, but tokens of pain are token identical with certain physical properties. In any event, both views require nothing more than physical ontology to secure the pain facts (modulo the phenomenal aspects).
naturalization strategy will also be similar. In the ethical case, we look for those properties that best fit the roles articulated by the system of platitudes. As with psychology, ethics will be naturalized to the extent that the realizer properties are natural properties.

Ralph Wedgwood (2001) has offered another variant on role semantics. On his view, the semantic roles of ethical terms are defined not by their place in a system of platitudes, but by their place in systems of sentences licensed by rules of practical reasoning. To figure out whether ethical properties are naturalizable one discovers whether the entities, properties or relations assigned as referents for ethical terms (where we assign referents to make the system of ethical sentences come out correct) would be naturally kosher. If so, ethics requires no additional ontology above and beyond the natural.

These views promise to achieve naturalization by connecting ethical terms with natural, in-the-world properties sans synonymy. In what follows I want to argue that a keystone normative term—‘reason’—does not refer to natural properties, and so if there are reason facts, they are non-natural facts, requiring an ontology above and beyond the natural. Later on I will have more to say that connects up ethical terms like ‘good’ and ‘right,’ and epistemic terms like ‘justified’ and ‘evidence,’ with reasons. To preview, the view is that various moral and epistemic claims are reason-implicating, and so they will inherit any of the metaphysical difficulties found in normative reasons quite generally. Indeed, all normative disciplines have metaphysical difficulties only insofar as they are reason-implicating.

B. A Minimal Normative Reductive Thesis

I want to articulate a minimal ethical naturalism that is consistent with the above semantic theories before I formulate a similar naturalism for normativity more broadly. There is an initial difficulty in characterizing \textit{natural} entities, properties, and relations, as opposed to
entities, properties and relations of other kinds. I will take the view that they are the properties picked out and described by idealized versions of the sciences.\textsuperscript{8}

A minimal ethical naturalism would not require anything like a conceptually grounded \textit{a priori} definitions (if and only ifs) of ethical terms using natural definiens. That is too close to the closed questions that Moore required. We need something weaker that still captures the naturalist’s thought that natural entities, properties, and relations in our world \textit{fix} all of the entities, properties and relations that there are in this world. As Kripke might say, once God added the natural facts, he did not need to take the further step of adding the ethical facts. On this view, any exact duplicate of our world with respect to the natural facts, \textit{and only the natural facts}, will also be an exact duplicate with respect to the ethical facts of our world. By duplicating the natural entities, properties, and relations of the world, we also thereby duplicate the ethical entities, properties and relations of the world. We can put this in terms of the following minimal ethical naturalism thesis:

\textbf{MEN:} Any minimal natural duplicate of our world is an ethical duplicate of our world.

If MEN in false, then either our world does not contain ethical facts, or it contains facts above and beyond the natural ones.

I want to concern myself with a view that is broader than MEN in two respects. First, I include in the minimal duplicate not just natural facts, but all of the non-normative facts,\textsuperscript{8}

\textsuperscript{8} Accord Timmons (1999, 12-13), Shaffer-Landau (2003, 59), Jackson (1998, 6-8). It seems clear that some other characterizations of natural properties—that they are the ones needed to explain structure and function, for example—will amount to the same thing. To be clear, I do not here argue for physicalism, or the view that all properties in the world are at base the properties studied by ideal physics. That is a view of how the special sciences relate to physics, and I am concerned with how ethics and other reason-implicating disciplines relate to the sciences, both special and basic. See Moore (1942) for his self-flagellation about his early characterization of natural objects and properties in terms of things existing in time.
including phenomenal facts, if such there be. I include phenomenal facts because some of them are clearly relevant to the normative status of things. Some acts are wrong, for instance, because they produce the qualitative feel of pain in others. If phenomenal facts are already included as kinds of natural facts, as naturalists would claim, then the explicit inclusion of phenomenal facts is superfluous, but harmless. Second, I broaden our inquiry and ask whether normative facts quite generally are duplicated when we duplicate the natural and phenomenal facts. For facts about whether one should pursue self-interest or the moral life, or whether one has reason to believe in the external, material world bear the marks of normativity, and it is an interesting, deep question whether such facts are duplicated along with the natural-phenomenal world.

With these modifications, here is the minimal normative reductive thesis, where the reduction base is the natural-phenomenal world:

MNR: Any minimal natural-phenomenal duplicate of our world is a normative duplicate of our world.

Notice that this is a metaphysical thesis. It is not a strictly reductive view if we take reduction to require an elimination of the concepts and explanations of the special sciences in favor of the concepts and explanations of basic physics. It is also not strictly reductive if that requires identities between the natural kinds and laws of the special sciences and the natural kinds and laws in physics. We want to respect the autonomy of the special sciences, social sciences, and all other domains and simply ask whether those facts are fixed by the natural-phenomenal world, or whether they require additional ontology.

---

9 In philosophy of mind, sometimes the natural world is meant to include both the physical and phenomenal ways things are. This is not traditional in metaethics, which is a literature I want to engage. Thus, I will use natural as it is often used in metaethics to refer to the ontology posited by ideal scientific inquiry.
The minimality requirement is important. Frank Jackson (1998) includes it in his own physicalism principle. Here is what he says:

[A] minimal physical duplicate of our world is a world that (a) is exactly like our world in every physical respect . . . and (b) contains nothing else in the sense of nothing more by the way of kinds or particulars than it must to satisfy (a). Clause (b) is a ‘no gratuitous additions’ or ‘stop’ clause (13).

Let me emphasize one important aspect of the minimality requirement as it relates to normative reduction. Suppose that there are strong metaphysical links between natural-phenomenal properties and some other kinds of properties such that it is metaphysically impossible to duplicate the natural facts of the world without bringing non-natural facts in train. For example, one might not be metaphysically able to duplicate the natural-phenomenal properties of the world without getting immaterial and undetectable souls for free. This would be a bit like choosing a person to marry and getting the in-laws for free. You might prefer to establish only the marriage relations with your beloved, and you might happily do without the in-law relations, but you simply cannot establish marriage relations with this person without also bringing some particular in-law relations in train. It can’t be done. This analogy only hints at the issue. We should imagine a more extreme case where all of the natural-phenomenal properties of this world are metaphysically necessarily towing along some distinct kind of property. Clearly, a world in which the natural-phenomenal properties are metaphysically towing along immaterial souls, or more relevant for our case, ethical or normative properties, is a world that contains more than the natural and phenomenal. Thus, we should read our minimality requirement as an attempt to preclude brute metaphysical necessitation between distinct kinds of properties, including strong supervenience relations between distinct ontologies.10

---

10 As far as metaphysical reduction is concerned, MNR is consistent with so-called non-reductive naturalism, whereby higher-order properties that are realized by natural-phenomenal properties
My project in the following, section II, is to argue that MNR is false with respect to reasons. As a result, if our world contains reason facts, it is not a minimal normative duplicate of itself.

II. The Unnatural Reason

I want to be as clear as I can about what reasons are. I am not speaking of motivating reasons or explanatory reasons, but of normative reasons. Motivating and explanatory reasons help us to explain and predict the behaviors of various agents. We might want to know why Sally killed Billy, for example, and a perfectly valid response would be: “The reason Sally killed Billy was to avenge her sister’s death.” Revenge is a nice motivating reason that explains why Sally did what she did. But there is another, distinctively normative question whether Sally had a good reason for killing Billy. The fact that Sally had a motivating reason to kill Billy does not entail that Sally had a good normative reason to kill Billy. Having said that, I am not ruling out the possibility that all of one’s reasons are in some sense connected to one’s motivational states. I merely want to point out the conceptual difference between explanatory reasons and normative reasons.

As a conceptual matter, the best explication of a normative reason is in terms of a favoring relation\(^\text{11}\) that holds between some reason-providing consideration and the action of some agent in some particular circumstance.\(^\text{12}\) Symbolically, we can represent reason claims as the following four-place predicate:

\(\text{favoring relation}^{11}\) that holds between some reason-providing consideration and the action of some agent in some particular circumstance.\(^{12}\) Symbolically, we can represent reason claims as

\[\text{in this world are realized by non-natural-phenomenal properties in other worlds. MNR is satisfied so long as 1) one can run so-called token-token identities between the natural-phenomenal realizers and instances of normative properties in our world, and 2) the same token-token identities hold in all minimal natural-phenomenal duplicates of our world (i.e., so long as the normative strongly globally supervenes on the natural).}\]

\(^{11}\) There will be disfavoring relations in the case of reasons against actions.

Read ‘F is a reason for A to Φ (or counts in favor of A’s Φing) in circumstances C,’ where, in the practical case, F is the favoring consideration, and A is the agent who can perform some act Φ in circumstances C. To illuminate the concepts involved, let us consider a concrete case. Suppose you are on the subway when the person behind you says “Ouch! You are standing on my foot.” You look back and notice that you are indeed stepping on a woman’s foot. In this case I would say that the fact that you are causing this woman pain (F) counts in favor of you (A) stepping off of her foot (Φ) given, inter alia, that you can easily do so without causing anything else bad to happen (C). If you are an ethical person, you will also take this consideration to count in favor of stepping off of her foot.

Colloquially, we often call the favoring consideration F ‘the reason,’ as we might say that the fact you are causing her pain standing where you are (F) is a reason to stop standing where you are. Now, that you are causing her pain standing where you are (F) is a fact of the natural-phenomenal world. So under colloquial usage it might look like this particular reason is found in the natural-phenomenal world. But calling F ‘a reason’ is an elliptical way of referring to the fact that F counts for or against some action or attitude. Crucially, locating the relata expressed by F in the natural-phenomenal world is not enough to locate the way in which it is supposed to normative relate to our actions and attitudes, either by favoring or disfavoring them.\(^{13}\)

In keeping with common usage, I will call the various Fs “favorers,” “reason-providers,” or simply “reasons,” understanding that the crucial aspect of the claim concerns its relation as either favoring or disfavoring some action or attitude, and I will call the whole relation itself

\(^{13}\) I think that David McNaughton and Piers Rawling (2003, 31) and Derek Parfit (1997, 124) are gesturing at this favoring relation when they claim that facts about considerations that “give” reasons are irreducible facts.
either a “reason,” or a “favoring relation.” I hope this somewhat casual use of the terms won’t cause confusion.

Thus understood, not everyone will agree that you have a reason to step off of the woman’s foot in our example. Nonetheless, the general commitment to normative reasons abounds. Many believe that agents necessarily have reasons to pursue the means to their contingent ends, again where this is understood as something more than a mere motivation to pursue their ends. To make a genuine normative reason claim, the thought must be that the fact that some act would advance a given agent’s ends counts in favor of that agent so acting, even if you cannot convince or motivate the agent to do it. Reason statements also abound in epistemology. We (or at least some of us) take ourselves to have reasons to believe in the entities, properties and relations posited by our best physics, and more mundanely, we have reasons to believe in the external world, that we have two hands, that we are not brains in vats, etc. Even in cases where individuals are not “epistemically motivated” to form beliefs in accordance with the evidence available to them, we rightly maintain that the evidence available to them provides them with reasons.  

A. A Test for Minimal Normative Reduction

Under minimal normative reduction, MNR, we must determine whether a minimal but complete natural-phenomenal duplicate of our world necessarily duplicates all of the normative facts of our world (again emphasizing that the minimality condition precludes brute metaphysical connections between distinct kinds of properties). To decide this issue, I suggest that we employ

---

14 As I see things, in the epistemic cases my reason providers—the things that favor my forming some beliefs and not others—are actually the contents of my mental states. For instance, the content of my appearance state of being appeared to redly and apply shapedly is evidence that, and a reason to believe that, there is an apple before me. But the particulars of various theories need not concern us. What is important is that most everyone thinks there are reasons in a variety of domains.
a test often used in metaphysics, and particularly in the philosophy of mind (see, e.g., Horgan 1984, 2006, Chalmers 1996, Jackson 1998, Chalmers and Jackson 2001, Lewis 1994),\(^{15}\) whereby one reduces a certain set of properties H to some other set of properties L by showing that there is an \textit{a priori} entailment from the L truths (and nothing more) to the H truths, or perhaps more precisely, there is a \textit{deduction} from the L truths (and nothing more) to the H truths.\(^{16}\)

Let me say a few words to clarify exactly how I understand this deduction test, and motivate its use. Consider the question of whether the water truths of our world require any ontology above and beyond the micro-physical facts of our world (as it now seems clear they do not). Following Terry Horgan (1984, 2006), we can approach this question by considering a super-agent S of superhuman cognitive abilities such that he can hold unlimited amounts of information in his head, and flawlessly reason through the consequences of his beliefs. Suppose that S knows all of the micro-physical truths of our world, and that S has complete conceptual mastery of the concept ‘water’ (which might include a grasp of its functional role, and thereby include conceptual mastery of other terms). Given the micro-physical truths, can S come to know the water truths? It would seem so, for S merely has to apply his concept ‘water’ to the micro-physical truths. Given conceptual competence, S knows that water is whatever it is that is that is

\(^{15}\) C.f. Block & Stalnaker 1999.

\(^{16}\) Chalmers and Jackson (2001) at times prefer that the test refer to “\textit{a priori} reasoning rather than explicit deduction” (338). I think the reference to \textit{a priori} reasoning only obfuscates the issues. It is clear that the information in the antecedent of the inference is taken as given, and need not be known \textit{a priori}, or even justifiably believed for that matter. Calling the resulting inference to the reducible truths an instance of \textit{a priori} reasoning leaves it unclear exactly what form the inference takes, whether it be deductive, inference to the best explanation, or what have you. Chalmers and Jackson think that some inference patterns provide at best brute, epistemically primitive metaphysical or nomological laws, and not property reductions, so it is crucial to understand exactly what form the reducing inference takes. It would appear that inference to the best explanation, and indeed all non-deductive forms of inference would fail to generate property reductions. Deduction appears to be the only inference pattern that can supply the desired ontological identities, but at times Chalmers and Jackson are not as clear on this point as one would hope.
clear, potable, falls from the sky, etc., and given the micro-physical truths S knows that \( \text{H}_2\text{O} \) is the clear, potable liquid that falls from the sky, etc. We can think of the deduction test as a way of capturing S’s ability to apply his concept of water to the micro-physical world, and we can represent S’s reasoning with the following conditional that can enter into a valid deductive syllogism:

If the micro-physical truths are thus and so, then the water truths are such and such.

The deduction test is clearly sufficient for the kind of minimal ontological reduction we are after. If the water truths are deducible, then no ontology above and beyond the micro-physical is needed to secure these truths. Whether deduction is necessary for reduction is a more difficult question. But consider what it would be like if our super-agent S had complete conceptual mastery of ‘water,’ was provided all of the micro-physical truths, and yet could not apply his water concept to those truths to determine the water truths. If S knows all of the micro-physical truths and yet cannot apply his concept of water to those truths, then he either does not yet grasp the concept of water (contrary to our hypothesis), or he is missing some crucial bit of information not contained in the microphysical truths, information that need to fix the water truths, and information that will require ontology above and beyond the micro-physical.\(^{17,18}\)

---

\(^{17}\) An important possibility is that the truths at the higher level will have to be fixed en bloc. If the semantics of some set of higher level truths are inter-dependent, as with functional role accounts, our super-agent S might have to apply all the concepts at the higher level at once, holistically, to the lower level truths. In this case, the set of higher-level truths will be deducible from the complete set of lower level truths. Also, for some cases, it might not be possible to go from the micro-physical truths to some higher level truths directly. Given the micro-physical truths the super-agent might not be able to directly infer truths about monetary transactions, for instance. S might need intervening conceptual levels, building up from physical truths to chemical and biological truths, and finally to complex social transactions.

\(^{18}\) Aside from these questions of ontology, there are additional modal questions about whether the \( \text{H}_2\text{O} \) facts necessitate (in some sense) the water facts. I address these questions, and their relation to our ontological questions, in section VI.
Because I will be considering a normative relational property, let me consider a more analogous non-normative case: the reduction of the relational property \(<\text{taller than}\>). Are the truths about who is taller than whom fixed by the height truths? Under the test, we ask whether our super-agent S, armed with the concept of taller than, and given all and only the height truths can deduce the taller than truths. He can, again by applying the concept of taller than. If he cannot, then he either does not grasp what ‘taller than’ means, or he lacks crucial information not found in the height truths.

Of course, the hardest case to date is the question of whether the phenomenal facts require facts above and beyond the physical facts. It looks like the phenomenal truths—truths about how one is being appeared to—are not deducible from the physical truths alone. Having the complete set of physical truths of our world, our super-agent could not yet determine whether this is a zombie world or a world of experiences simply by applying concepts of phenomenal properties. To determine that the super-agent must have access to more information regarding the what-it’s-likeness of experiences in the actual world, and which physical properties these experiences supervene upon. In short, he must have information about facts that outstrip the physical facts, or so the dualist can argue.

Of course, there are some potential rejoinders to this deductive failure, and after setting out the case for normative non-reduction I will consider whether those strategies can be made to work for the normative case. But beyond these brief remarks, and some clarifications in the sections below, I will not defend the deduction test at length here. Suffice it to say that if this argumentative structure works in metaphysics and the philosophy of mind, it will work equally well in the case of reasons, as the following arguments show.
B. The Reason Gap

Recall once more minimal normative reduction thesis:

MNR: Any minimal natural-phenomenal duplicate of our world is a normative duplicate of our world.

Applying the deduction test, MNR is true if and only if the normative truths are (in principle) deducible from the natural, phenomenal truths and nothing more. The question I will focus on is whether the reason truths are deducible from the natural and phenomenal truths (NP truths) and nothing more.\textsuperscript{19} If not, then MNR is false.\textsuperscript{20} To focus on the in-principle question we can use our heuristic device and ask ourselves whether a super-agent S can determine the reason truths simply by applying his normative concepts to the NP truths when given only 1) complete conceptual mastery of the normative and 2) the complete and fully discriminating set of NP truths. If not, it looks like the NP truths do not fix the normative truths after all, for our agent needs more information before he can apply his concepts one way or the other.

Let me first consider our example where you are causing a woman pain by standing on her foot, and where you can relieve that pain by stepping off her foot without causing anything else bad to happen. Imagine a complete description of the situation in NP vocabulary and ask

\textsuperscript{19} Unfortunately, there is a snag if we only work with natural and phenomenal truths. Consider a statement like “Those things over there are blueberries.” If we only know the natural and phenomenal truths, we won’t know whether this statement is true because we won’t know who is uttering it and where “over there” is. To be able to deduce truths that include indexicals we must also include indexical information along with natural and phenomenological facts. Though it is sometimes thought to be significant regarding consciousness, this is a minor point and will have no significance in the normative domain. But to be absolutely precise, the question is whether the normative truths can be deduced from the natural, phenomenal and indexical (NPI) truths.

\textsuperscript{20} This test does not revert to the old conceptual analyses that Moore desired. I make no assumptions that the higher order concepts can be fully analyzed using the lower level concepts. All the test requires is a conditional claim: if the lower level truths are \textit{thus and such}, then the higher level truths are \textit{this and that}. Thus we see whether the natural-phenomenal truths are sufficient for the normative truths, but leave open what conditions, if any, are necessary for the normative truths.
whether the reason truth, if such there be, is deducible. I am on record: I think that the fact that you can relieve this woman’s pain is a reason for you to step off her foot regardless of whether you care to relieve her pain. But our question is not whether any particular person would make such a reason judgment. Indeed, those more fond of end-based reasons would judge that you do not have a reason to step off of her foot unless it would advance some contingent end of yours. Our differing reason judgments aside, to consider the question clearly we should imagine a super-agent S with complete knowledge of the NP truths and complete conceptual competence with respect to reasons, and ask whether he can deduce the reason truth using those resources alone.

The problem is that those resources are too impoverished to enable the deduction. Our super-agent might know, for instance, that that stepping off the woman’s foot would relieve her pain, and doing so would not further any contingent desires of yours. But those facts are not yet reason facts, for they say nothing about whether such considerations count as reasons for action or not. Our super-agent might also grasp the concept of a normative reason to Φ: it is a consideration that counts in favor of Φing. Yet armed with all that, he still would not know exactly what favors what. As far as the NP truths and the concept of a reason are concerned, it is entirely open whether reducing the pain of another counts in favor of action, or whether only personal end satisfaction of some sort counts in favor of action.

Of course, it is difficult to imagine an agent who grasps the concept of a reason that does not also have some additional normative commitments concerning what favors what. If we supplement our agent’s resources to include not just the NP facts and the concept of a reason, and allow him to employ some additional commitments about what favors what, then he might deduce some conclusion about whether you have reason to step off the woman’s foot. If the
supplemental commitments include the belief that reducing the pain of others favors action, then he will conclude that you do have a reason to step off her foot. If the supplemental commitments include the belief that only your end satisfactions favor your actions, then he will conclude that you do not have a reason to step off her foot.

But notice that allowing this super-agent some additional normative commitments to secure the deduction will not suffice to reduce reasons to the NP world. That would be like locating phenomenology in the natural world by deducing phenomenal truths from the natural truths, conceptual competence with respect to phenomenal concepts, and information about what it’s like to experience this world. If we make use of information about what it’s actually like to experience the world, information not contained in the natural facts or phenomenal concepts, we fail to show that the phenomenal facts require no ontology above and beyond the natural facts. Similarly, to ensure that reason properties are locatable in the natural-phenomenal world, there must be a deduction that uses only two resources: 1) the NP truths, and 2) conceptual competence with respect to reasons. If extra information is needed to secure the deduction, it looks like it will be information not already included in the NP truths, thus requiring some facts, and so some ontology, above and beyond the NP ontology.

C. Super Disagreement

How can we be so sure that some more determinate conception of reasons—instrumentalism, internalism, or externalism—is not found in the very concept of a reason? How can we be so sure that we must reach outside of the concept to find these extra commitments I speak of? Perhaps, one might claim, the concept of a reason is not just the concept of a consideration favoring action, but it is the concept of a particular consideration, desire
satisfaction, favoring action. If the concept of a reason is this robust, then perhaps the super-agent can apply it to the NP truths after all to deduce the reason truths.

The problem with this line of argument is that it incorrectly rules out certain scenarios as conceptually impossible. On this line of argument, any two super-agents, S1 and S2, given NP truths and conceptual competence with respect to reasons, and who apply those concepts perfectly (including non-acceptance of contradictory beliefs), would have to come to the same conclusions. It would be *impossible* for S1 to conclude that you do have a reason to step off the woman’s foot, and S2 to conclude that you do not. Yet surely this is possible. S1 and S2 might wholly agree on all the NP truths, and yet disagree about what you have reason to do. S1, for instance, might think that pain relief counts in favor of stepping off her foot. S2, on the other hand, might think that only desire satisfaction counts in favor of any actions. Nothing in NP settles this dispute, and it is too high handed and desperate to suggest that one of these super-agents does not grasp, and so does not employ, the concept of a reason. Only as a last resort do we settle substantive normative disputes like these by charging opponents with conceptual confusion.

This kind of disagreement about what one has reason to do is conceptually possible. S1 and S2 grasp the very concept of a reason, they simply disagree about what favors what. Again, what would generate the disagreement is nothing found in NP or the concept of a reason, but additional normative commitments that these super-agents have about what considerations actually do favor actions and attitudes, commitments that are not entailed by the bare concept of a reason. So if one of these agents judges truly and the other falsely about whether you have reason to step off the woman’s foot, there must be some facts above and beyond the NP facts that renders things so, viz., facts about what favors what.
D. Fragmenting the Concept?

S1, we have imagined, believes that agents only have instrumental reasons to act. S2, we have imagined, believes that agents have basic reason to relive the pain of others. What if we simply hold that S1 and S2 are applying different concepts of different kinds of reasons? S1 is applying the concept of an instrumental, hypothetical reason, or reason\(_{hi}\). S2 is applying the concept of an external-categorical reason, or reason\(_{ci}\). So S1 looks at all the NP truths and applies the reason\(_{hi}\) concept to deduce that you do not have reason\(_{hi}\) to step off the woman’s foot. S2 looks at the very same NP truths, but applies a very different concept, reason\(_{ci}\), to deduce that you do have reason\(_{ci}\) to step off the woman’s foot.

This strategy assumes that there are both hypothetical and categorical reasons for action, and it tries to fit both in the natural world by building in the considerations from which favoring flows into different reason concepts, which, when applied to NP, deliver reason truths. One problem with the strategy is that it invites embarrassing riches of reasons, for if we can build in considerations from which favoring flows into different reason concepts, we can create all kinds of reasons for action by adopting the respective notion of a reason. To locate in the NP world reasons to harm others for fun, for instance, then apply the concept of reason-for-harming-others-for-fun, according to which the harm caused to others is a reason to perform the harmful act. This is too cheap. If there are certain properties in the world, including the favoring relations found in reasons, we must look to the world to find them, and we cannot include them in our ontology through conceptual fiat.

The more significant problem with this strategy is that it fails to capture genuine normative disagreements about what we have reason to do. In our scenario, S1 thinks that you only have reason to pursue your contingent desires, and S2 thinks you have reason to do things
that might not count as pursuit of your contingent desires. But if these two are simply employing different concepts, then they are not having the genuine disagreement that they think they are having. Because they are employing different concepts, they are speaking of entirely different subjects. They are talking past one another. Their “disagreement” would be no more a disagreement than the “disagreement” between two agents about whether a bank is the best place to keep money, where one agent is talking about storing money in the side of a river, and the other is talking about storing money at a federally insured fiscal institution.

To preserve genuine disagreement between instrumentalists, internalists, externalist, and other reason theorists, they have to be talking about the same thing. Fragmenting concepts of reasons ensures that they are talking about different things. Instead, we should take reason disagreements at face value: Humeans, Kantians, Utilitarians and others are arguing about what favors what. And if that is the case, then whatever favoring truths hold in this world, they are not deducible from the concept of a reason and the NP truths alone. Hence, favoring facts require ontologies above and beyond the NP ones.

E. Epistemic Reasons

So far I have focused on locating practical reasons in the natural-phenomenal world. As we have seen, even instrumental reasons cannot be located because they, too, are committed to the favoring relation found in all reasons. An even larger pill to swallow is that epistemic reasons cannot be located in the natural world. Here I want to apply the same strategies above to show that there is no entailment from the NP truths to truths about epistemic reasons (of course, employing the concept of a reason).

The structure of epistemic reason claims is similar to all reason claims, but in the case of epistemic reasons we are talking about considerations that favor belief in various propositions.
As in the practical case, these favoring facts do not show up at the level of NP, nor are they built into the concept of a reason such that one armed with all the NP truths can simply employ the concept of a reason to deduce the epistemic reason truths.

Nonetheless, the possibility of some bridge premise spanning NP and epistemic reasons looks more promising than bridge premises in the practical case. The best place to start is truth. It is hard to deny that one has reason to believe the truth. In fact, some kind of truth connection seems to distinguish epistemic reasons for belief from practical reasons for belief. And yet there are telling cases where we have epistemic reasons to believe propositions that are clearly false.

Take a case where an agent, Andy, is abducted in the dark of night. Still in bed, he is placed in a fun house that is full of mirrors and deceptive illusions. When Andy wakes up he has no idea where he is, but it appears to him that there is a sharp abyss to the right of his bed, and a sturdy floor extending to the left of his bed. In fact, the opposite is the case. There is an abyss to the left of his bed, but a sturdy floor to the right, though the fun house is set up so that Andy thinks otherwise. In this case, does Andy have reason to believe that there is an abyss to the right of his bed or the left? I would say that Andy has epistemic reason to believe what he sees, so he has epistemic reason to believe that the abyss is on his right, as it appears to be. If he believed otherwise, he would fail to align his beliefs with his evidence.

Here is a case where an agent has epistemic reason to believe something false, and has no epistemic reason to believe what is true. From the fact that some belief is true, nothing follows about what one has reason to believe. There might, however, be a more tangential relationship between epistemic reasons and truth. What if we say that Andy’s information renders his (mistaken) beliefs about the location of the abyss likely true? Perhaps we can cash out likely truth through some notion of process reliabilism, or via some other tools available in the NP
world. Assuming *arguendo* that some such notion of likely truth can be made to work, have we located epistemic reason facts in the NP world?

I think not. For even if we know that some belief is likely true, there is still an open question whether one has reason to accept beliefs that are likely true. While facts about the likelihood of truth might show up at the level of NP, no facts about reasons show up at the level of NP; we still need to bridge the gap between what beliefs are likely true and what propositions we have reason to believe. Of course, one could argue that applying the following conditional is constitutive of the concept of an epistemic reason:

Likely truth: If p is likely true given A’s information, then A has reason to believe that p. This conditional holds, it would bridge the gap between the NP facts found in the antecedent, and the reason facts found in the consequent, thereby ensuring that no extra ontology is needed for the reason facts. But how plausible is it that this conditional is constitutive of epistemic reasons? If someone denies it, does she thereby fail to grasp the concept of an epistemic reason? It seems not. Susan Haack, for instance, denies that truth has anything to do with evidence and justified belief: “[A] reliabilist account of justification is just incorrect. Justification is a matter of the *experiential anchoring* and *explanatory integration* of the subject’s *evidence* with respect to a belief; and explanation in terms of the *truth-conduciveness* of *belief-forming processes* simply uses the wrong concepts and consistently yields counter-intuitive consequences” (Haack 1993, p. 139, emphasis in original). And we can imagine a very sophisticated religious believer who acknowledges that his belief in God is not likely true, and yet maintains that he has reason to believe in God because, for instance, an authoritative religious text tells him to.

More to the point, we can imagine two super-agents, S1 and S2 who know all of the NP truths, including truths about what beliefs are likely true based on certain information, and yet
who disagree about what a particular agent has reason to believe. Perhaps S1 is a committed Berkelean. He is given all the information in NP, knows how an agent, Andy, is appeared to, etc., and maintains that Andy does not have reason to believe in the external, material world, even though it is *highly likely* (understood naturalistically somehow) that the material world exists. S2, on the other hand, maintains that Andy has reason to believe in the material world, including beliefs that he has two hands, that the abyss is on his right, etc. In this case, neither agent lacks the very concept of an epistemic reason for belief. They both apply that concept to the information in NP, but they supplement their resources with additional commitments about what considerations epistemically favor belief formation. These extra commitments generate their disagreement; the commitments are not found in NP, nor are they essential to the concept of an epistemic reason. If one super-agent has *correct* extra commitments about what favors what, there must be facts above and beyond NP that makes these commitments correct. In short, epistemic reasons are not deducible from NP and the conceptual competence with respect to epistemic normativity alone, thus epistemic reasons relations have not been located in the natural-phenomenal world.

It is worth emphasizing that I am not arguing that *no* normative facts can be located in the NP world. But I do consider reason facts to be the *sin qua none* of normative facts, the holy grail of normativity, if you will, and I think that no reason facts, and so no reason-implicating facts, can be located in the natural world. In the epistemic cases, even if there are aspects of justification or knowledge that require no extra ontology above and beyond the NP ontology, these facts will not answer the most important epistemic question we can ask: What do we have reason to believe?
This concludes the basic case for the normative ontological gap. In what follows I want to further support the gap by addressing some objections couched as fallacious but tempting gap-bridging arguments.

III. Red Herrings

Notice again that I do not claim that no normativity can be located in the natural-phenomenal world; I only claim that the normativity found in reasons cannot be so located. However, there is a temptation to glide from facts in the NP world that bear some semblance of normativity to reason facts, so I want to spend some time fortifying the normative ontological gap between aspects of naturalized normativity and reasons.

A. Hypothetical Imperatives

Those who advocate hypothetical imperatives can be rather insistent that the reasons produced therefrom are naturalistically kosher. They often point to moral reasons unconnected with one’s contingent ends as the red-headed step-child of normative reasons. Hypothetical imperatives are usually couched at the overall level in terms of oughts, but couched at the level of reasons they can take the following form:

M-E principle: If means M is instrumental to achieving A’s end E, then A has reason to take means M.

In other words, the fact that M is instrumental to achieving some end E counts in favor of taking means M, though it can be outweighed by other reasons to take different means to this end, or to pursue other ends. Insofar as one is committed to reasons, however, one is committed to favoring relations and these are things that cannot be located in the natural-phenomenal world. This principle certainly has the right form for solving the location problem. If true, it would provide the bridge premise from NP truths to reason truths. But to solve the location problem we
must find this bridge premise in the NP world or in our conceptual competence with respect to reasons.

The principle is not found at the level of NP. While elements of the antecedent are certainly part of the NP world, the conditional itself is not. It purports to go from truths of the NP world, truths about desires and the consequences of various actions, to reason truths, but it itself is not found in the NP truths. So, is this principle part of conceptual competence with respect to reasons? Well, consider a super-agent who does not think one has basic reasons to pursue one’s ends, but does think one has basic reasons to help others in need, for instance. As above, the best description of this super-agent is not that he fails to grasp or employ the concept of a normative reason. Instead, he employs the concept of a normative reason supplemented by additional commitments about what favors what.

It is open to advocates of hypothetical imperatives to claim that there is nothing more to their view other than facts about ends and the consequences of means. In that case, the view becomes entirely non-normative and makes no commitments about what one has reason to do. To be sure, a super-agent armed with the NP truths and normative conceptual competence will know that M is a means to E, or that it will achieve E, and he might know that agent A believes that he has reason to take means M, for these truths show up at the level of NP. But it is still an open question whether A actually has reason to take means M. The non-normative view that eliminates reasons in favor of talk about what actions achieve which outcomes offers no guidance to those who wish to ask the gripping normative questions: What do I have reason to do? and What do I have reason to believe?

B. Ideal Rationality

---

21 I thank Dave Schmidtz for suggesting this helpful way of putting the point.
What if we embellish the crude version of the hypothetical imperative? Consider one such reasons internalist theory, according to which an agent A has reason to Φ just in case an ideally rational version of A would be motivated to Φ. In other words, the fact that an ideally rational version of yourself would be motivated to Φ counts in favor of you Φing. From this we might generate the needed bridge premise from the NP truths to some reason truths by taking the conditional that runs from ideal rationality to reasons:

Reasons Internalism: If A’s ideally rational self would be motivated to Φ, then A has reason to Φ.

We can secure the deduction so long as ideal rationality can be cashed out in NP terms.

As with crude instrumentalism we must also establish that some such bridge premise is deducible from NP alone, or is a matter of conceptual competence with respect to normativity. The first option is a dead end. Favoring relations are radically different kinds of things than anything else found in NP. It is hopeless to try to deduce truths about favoring at the level of NP. The only real option is to establish that some such bridge premise is a matter of conceptual competence. That is, anyone who does not use the term consistent with this principle does not really have the concept of a reason.

That sounds rather farfetched. For any natural description of an ideally rational agent that provides bridge premises from the natural to the favoring relations, one can agree that such a person would Φ, and yet deny that he has reason to Φ, all the while fully appreciating the concept of a reason. This is what enables those who believe that there are only instrumental reasons for action to honestly disagree with those who think we necessarily have reasons to be moral. And, considering full information accounts of rationality, Alan Gibbard (1990) considers a more general problem. He notes that a fully informed version of himself might come find digestion

disgusting: “Perhaps with a more vivid realization of what peoples’ innards are like, I would want to stay away from dinner parties and do all my eating alone—although then I would feel lonely and isolated” (20). He concludes that it does not follow that it would be rational, or recommended, that he never eat with other people. We can conclude that it does not follow from the fact that an ideally rational version of oneself would be motivated to $\Phi$ (naturalistically understood) that one actually has reason to $\Phi$.

If conceptual competence generates the reasons internalist premise, the reasons externalist would be conceptually incompetent, and there would be no genuine disagreement between the internalist and the externalist, just different people talking about different things. But why insist that disagreements here are the result of conceptual incompetence? It looks like we can all understand that reason claims are about what favors what, and we simply disagree about what favors what. This disagreement is not about features of the NP world, and it is not settled by the facts of the NP world.

The internalist principle might look like a matter of conceptual competence because it trades on the notion of rationality. For there is a relationship between reasons and rationality. However, the only conception of rationality that is related to reasons as a matter of conceptual competence is a hopelessly circular one, whereby a rational person is sensitive to reasons. Clearly this does not establish a principle that would bridge the gap between the NP truths and the reason truths, for now there are no sufficient conditions cashed out at the level of NP that would allow us to move from there to some conclusion about reasons. To get the bridge premise we must work with a notion of rationality that is found at the level of NP, and once we do that we find that the conditional is not found at the level of NP, nor is it a matter of conceptual
competence. For any naturalized description of a rational agent, there is an open question whether one has reason to do what an agent of that description is motivated to do.

Ultimately, it seems that the favoring relation found in reasons is a basic notion. We can understand various other concepts in terms of reasons, but conceptual competence with respect to ‘reason’ requires no more than the idea of favoring relations, and this favoring relation is so radically different in kind from anything found in NP that it looks like there will be no candidates in NP for realizing it.

C. Proper Function/Evolutionary Design

Evolutionary design is an attractive place to look for normativity in the natural world. If a bee is designed by nature to seek pollen from a certain flower, for example, and that bee fails to visit those flowers, we can say that the bee is not functioning properly. Proper functioning, as opposed to garden variety functioning, has a normative flavor to it. In the case of the bee, proper functioning is something determined by selective processes operating over reproductive generations of bees. And it is no stretch to assume that certain activities of human beings count as their proper functions from an evolutionary perspective, while certain activities might be considered malfunctions. From this modicum of normativity can we extract reason facts? That is, does the claim ‘A has reason to Φ’ follow from the fact that A’s Φing would be a proper functioning of A or A’s species?

Certainly not. To think that some reason conclusion follows from design or proper functioning is to commit one version of the naturalistic fallacy. Evolution can design beings to do all kinds of horrible things. Perhaps evolution designed us to be prejudiced against out-group members (i.e., those we do not fall within whatever social class we identify with). If so, it would not follow that we have reason to be prejudiced against out-group members. Perhaps, on the
other hand, our “selfish genes” guide us to engage in behaviors that benefit them, or some set of
them. If so, it would not follow that we have reason to engage in those behaviors for the benefit
of those genes. We can intelligibly ask whether we should not do something that benefits us, the
agents (either individually or collectively), rather than some proper subset of our genes. More
generally, for any behavior that we are designed to do, or which can be characterized as proper
functioning, there is an significant open question: do we have any reason to do what we were
designed to do? If it turns out that we do have reason to engage in some behaviors for which we
were designed, this fact would not show up at the level of NP; it would be an extra fact about the
world in addition to the natural facts about selective processes operating on reproductive
generations. Thus, no story about nature’s design or a resulting proper function serves to locate
reasons in the natural world.

D. Why we Are Mislead

All of the candidates we have considered offer perfectly natural properties as the
considerations that count in favor of or against various actions or attitudes. The instrumentalist
and internalist analyses, for example, suggest that reasons are given by the motivations of agents
described in natural terms. But the naturalizability of these favoring considerations should not
mislead us into thinking we have naturalized the favoring relation itself. Recall that I earlier
distinguished the favoring consideration F as just one relata in the favoring relation. I do not
deny that one can offer a description of a rational individual in natural-phenomenal terms, and I
do not deny that the considerations that motivate such a person are describable in natural-
phenomenal terms. I merely point out that that only speaks to one relata (the F relata) of the
favoring relation found in reasons. And one cannot reduce a relational property merely by
showing that one relata of the relation is reducible. It would be like trying to reduce the taller-
than relation by showing that one person has height, or reducing the believes-that relation by showing that the contents of attitudes are naturalizable.

What any theory of normative competence must preserve is the idea of a favoring relation, for that is the heart and soul of reasons, whether they are instrumental, ethical, epistemic, or what have you. In the philosophy of mind, dualists often stress a similar point. They say that a full array of mental concepts cannot ignore the what-it’s-likeness of experience, as purely functional accounts seem to do. With what-its-likeness firmly in mind, dualists argue that these special properties cause problems for reducing consciousness to the natural. I am claiming that there is an element in our normative thought and language that cannot be ignored. We think and talk about reasons and in doing so we think and talk about what favors what. But once we attend to this all-important normative relation we see that it cannot be deduced from any of the NP truths.

Having said the above, many reason claims strike me as obviously true. But they strike me as obviously true not because I am merely applying the concept of a reason to natural-phenomenological facts. They seem true because I employ some additional commitments about what favors what, commitments that are not found at the level of NP and not essential to the concept of a reason. All the failure of deduction shows is that reason facts would be non-natural and non-phenomenal facts. If the argumentative structure goes through, and there really are favoring relations corresponding to the reasons I find obvious, they would be a third irreducible kind of fact, and our world would fail to be a minimal natural-phenomenal duplicate of itself.

IV. Phenomenal Lessons

Above I tried to establish and fortify the normative ontological gap. I addressed some arguments that purport to show that normative deductions are forthcoming, but I have not yet
addressed arguments that grant me normative non-deducibility, and that attempt to generate reductions despite non-deducibility. Here I want to fill that argumentative lacuna.

Can one grant that the deduction fails in the case of reasons and still maintain that the normative requires no ontology above and beyond the natural and phenomenal? We have to be careful with this kind of strategy, for it can appear ad hoc. The deduction test seems to work in (just about?) every case of ontological reduction, and we should not let the elegant simplicity of naturalism tempt us into making an ad hoc exception for a certain class of properties. Having said that, some philosophers of mind try to show that failure of deduction does not demonstrate failure of reduction. They employ this strategy in an attempt to explain why the phenomenal is not a priori entailed by the physical despite what they take to be de facto phenomenal-physical identities, so it will be instructive to consider whether similar explanations can be employed in the normative case.

Hill and McLaughlin (1999) and Loar (1999) provide material to examine the phenomenal case. They begin by driving a wedge in the conceptual domain, holding that phenomenal concepts are radically different from physical concepts. For Hill and McLaughlin, phenomenal concepts, or what they call “sensory concepts,” are recognitional concepts, used to introspectively recognize self-presenting sensory states. Physical concepts, or what they call “theoretical concepts” are not recognition in this fashion, though I take it that some theoretical concepts might be non-recognitional ways of conceiving of sensory states. (See Hill and McLaughlin 1999, 448-49). Similarly, Loar claims that phenomenal concepts and recognitional, unlike physical concepts, which are theoretical. (Loar 1999, 467). If these philosophers are right, and if phenomenal properties are actually identical to certain physical properties, perhaps the extremely different in the ways in which phenomenal concepts and physical concepts
conceive of properties can account for any super agent’s failure to deduce the phenomenal truths from the physical truths. Note that, in the case of phenomenology, tight correlations between certain brain states and certain experienced phenomenal properties might be reason enough to posit the some identity claims so long as the deductive failure can be explained away as we have just suggested.

Let me assume *arguendo* that an explanation along these lines is convincing in the case of phenomenal-physical reductions, and ask the following question: If there are no deductions forthcoming for reason truths, can we adapt these explanations to explain away the deductive failure? Is there reason to hope that normative reasons require no more ontology than the natural and phenomenal? I think not. We would need some explanation for how normative reason concepts are conceiving of some relations in the NP world in an utterly different way than do the natural and phenomenal concepts. No such explanation is forthcoming. Firstly, unlike the phenomenal case, there are no candidate relational properties found in the NP world that tightly and uncontroversial correlate with reason relations. Even if we wanted to run property identities, we would need to find a suitable relation in the natural-phenomenal world. Secondly, there is only *ad hoc* reason to suppose that normative concepts conceive of natural properties in unique fashion. In the phenomenal case, by contrast, there is some independent plausibility to the claim that we recognize certain self-presenting states (the manifest in experience) via phenomenal concepts. So this kind of reply to the deduction test is a non-starter in the case of normative reasons.

Instead, normative naturalists must reject the deduction test in its entirety. The success of their view will hang on how well they discharge their burden of defending various minimal reduction theses *sans* deductive entailments. With others, I have tried to motivate the general
applicability of the deduction test. The burden now lies squarely on the shoulders of those who
would reject the test in its entirety. They cannot simply respond to the positive arguments for the
deduction test; they must replace that test with something equally successful.

V. The Objection from Strong Global Supervenience

In making my case I have been relying on analogous arguments in the philosophy of
mind. There is, however, an important disanalogy between the phenomenal and normative cases
that merits further attention. In the phenomenal case, one often argues that the following
scenario is conceivable: a world that is materially exactly like our own, but where the beings
have conscious experiences that differ from our own.23 Perhaps the beings are spectral invertors or
complete zombies, in which case there is nothing it is like to be them. If such a scenario is
conceivable, as it seems to be, one can argue along the following lines:

(1) It is conceivable that there be zombies
(2) If it is conceivable that there be zombies, it is metaphysically possible that
there be zombies.
(3) If it is metaphysically possible that there be zombies, then phenomenology is
nonphysical.
(4) Therefore, phenomenology is nonphysical.24

In the normative domain, the exactly analogous argument would fail. The analogous
argument would have to establish that the following world is conceivable: a world that is
naturally and phenomenologically exactly like our own, but where the normative facts differ
from the normative facts of our world. This scenario is not conceivable. As a conceptual matter,
the normative facts strongly supervene upon the non-normative facts, which, for our purposes,
just are the NP facts. According to the strong global supervenience thesis, a world that is
identical to our own with respect to all non-normative facts would necessarily be identical with

23 See, e.g., Chalmers 1996.
respect to all normative facts. No change in the latter without a change in the former. Thus, in the normative case, we cannot argue from the conceivability of normatively differing duplicates. Does this show that reason facts are reducible after all?

Frank Jackson (1998) thinks so, at least as applied to ethical properties. He supplements strong global supervenience of the ethical on the descriptive with a theory of properties, according to which necessarily coextensive properties are identical, to argue that ethical facts are in fact reducible to descriptive facts.25 If by ‘descriptive property’ Jackson simply means to include all properties but the normative ones, then his argument purports to show that normative properties can be identified with long disjunctions of non-normative properties.

Without going into the detail of Jackson’s argument, let me point out a general problem with his account, and any account that bases reduction on strong supervenience laws. Without deductions running from the non-normative facts to the normative facts, strong global supervenience and necessary co-occurrence of properties is compatible with brute metaphysical necessitation between non-normative and normative properties, where these properties remain distinctly different in kind. The deductions are needed to ensure that no additional ontology above and beyond the non-normative is introduced. If there are normative facts that respect strong global supervenience without the deduction based on NP and conceptual competence of

---

25 Accord McNaughton and Rawling 2004. One concern of mine about Jackson’s argument is that it runs the identities between ethical properties and descriptive properties, and I have to confess that I do not know what descriptive properties are. ‘Descriptive’ is most aptly predicated of linguistic items like statements or propositions. I think I know what it is for a claim or some other linguistic item to be descriptive – it has to purport to represent a way the world is, including the way it is with respect to properties. But I have no idea what it is for a property to be descriptive. Properties don’t purport to describe anything. Properties just are. Predicating ‘descriptive’ of properties sounds like a category mistake. I don’t know whether Jackson would include ectoplasmic and magical properties as descriptive properties. I don’t see why not, for these properties can make the world the way it is. Similarly, the reason giving relational property might also count as a descriptive property, for it too can make a world the way it is just as much as the taller than relation and the believes that relation.
reasons alone, there must be brute, metaphysically necessary bridging laws between the non-
normative and the normative such that the non-normative facts necessarily (sub specie
metaphysicae) bring the reason facts in train. That is, the entities, properties and relations of NP
necessarily tow the reason relations along with them. This is an anathema to minimal versions of
naturalism, or natural-phenomenalism, as (I think) Jackson would agree.

While strong global supervenience blocks traditional formulations of the conceivability
argument, the point of the conceivability argument in the philosophy of mind is to show that
there is no a priori entailment from the natural truths to the phenomenal truths. To make that
basic point, one does not need to appeal to the conceivability of natural world duplicates that
have different phenomenal truths from those that hold in our world. All one needs to show is
that, given only the natural truths of this world, any number of phenomenal truths for this world
are conceivable.

Despite strong global supervenience of reason truths on the non-reason truths, we can
make the same point concerning reason truths of this world. Crucially, strong global
supervenience does not tell us which normative facts hold in two NP duplicates. It only tells us
that the same normative facts—whatever they happen to be—hold in the two NP duplicates. So
we retain the basic point of conceivability arguments if we home in on our world described only
in terms of NP, and ask whether there are a number of normative truths that are conceptually
possible for this world, but logically incompatible with one another. Given the NP truths of our
world, for example, the following two truths are logically possible but incompatible with one
another: 1) there is reason for progressive taxation, 2) there is no reason for progressive taxation.
Only one of these statements can be true in this world, and so, per strong global supervenience,
true in any minimal NP duplicate of this world. Yet neither we nor a super-agent S can figure
out whether (1) or (2) holds from NP and normative conceptual competence alone. Hence we argue from the fact that various incompatible normative truths are conceivable for this world, given NP, to the conclusion that normative facts cannot be reduced to NP facts. We need not go through any premise about the metaphysical possibility of NP duplicates having normative differences. Such premises are not essential to the basic structure of the argument.

VI. Other Matters

Before concluding, I want to address a couple of matters that will help to clarify my position and that will help to locate it in argumentative space with related issues. Below I consider modal truths akin to ‘water is necessarily H$_2$O,’ and Hume on the is-ought gap.

A. A Posteriori Metaphysical Necessities

One might wonder how, exactly, this deduction test relates to so-called a posteriori necessities, such as those found in the proposition “water is H2O.” As I use it, the deduction test has a very limited scope. It only determines whether any higher-level truth of this world requires properties above and beyond those needed for some lower-level set of truths of this world. Applying the test in the case of water, the question is not a modal one. It is a question of how water gets realized in this world. More specifically, do the water truths of this world require any ontology above and beyond the actual world ontology introduced by physics and chemistry? Under the test we ask whether water truths are deducible from only 1) the complete set of physical and chemical truths, and 2) conceptual competence with respect to ‘water.’ In this case, the answer is “yes.” Given all the truths about physics and chemistry, including the truth that H$_2$O is the clear, potable stuff that fills our lakes and streams, falls from the sky, etc., a super-

---

26 If some reason statement strikes you as clearly, obviously true, you must be careful not to conclude on that basis that it’s negation is inconceivable. A proposition can strike us as obvious and at the same time it can be conceptually possible that the proposition not obtain just in case no contradiction is involved.
agent S can employ his conceptual competence of water to determine all the water truths. The following conditional reconstructs his conceptual competence: if the watery stuff in this world is H₂O, then the water truths of this world are ________. There is no information that S is missing, and that is needed for him to determine the water truths. If there were, water truths would require truths, and so ontologies, above and beyond the physical and chemical ones.

This deduction test goes through whether or not our term ‘water’ necessarily picks out H₂O in all possible worlds (i.e., whether or not ‘water’ is a rigid designator of the stuff in this world that realizes the watery role). Consider two possible worlds, w and w*, that are in all respects identical to our own except for the fact that in one world, w, ‘water’ is a rigid designator, and in the other world, w*, ‘water*’ is not a rigid designator – in w* ‘water*’ simply picks out the watery functional role. In w, the H₂O truths are sufficient to fix the water truths, and a super-agent could apply his water concept to the chemical and physical truths of that world to determine the water truths. The following conditional holds: if the watery stuff in w is H₂O, then the water truths of w are ________. But likewise, in w*, the H₂O truths are sufficient to fix the water* truths, and a super-agent could apply his water* concept to the chemical and physical truths of that world to determine the water* truths. The very same conditional found in w holds in w*: if the watery stuff in w* is H₂O, then the water truths of w* are ________. So whether or not our term ‘water’ is a rigid designator or not, the chemical and physical truths of this world are sufficient to fix the water truths of this world. Rigid designation only shows up as a modal difference, concerning whether, in other possible worlds, our term ‘water’ picks out not H₂O, but whatever it is that fulfills the water-type functional role, if anything. These questions about how our term ‘water’ applies to other possibilities is not within the purview of the deduction test as we are using it. We only want to know whether water truths in
**this world** require more by way of ontology than physics and chemistry. The answer to that question is “no, physics and chemistry are sufficient to fix the water truths of this world.”

Not so for reasons. I have tried to show that reason truths are not deducible from the NP world. Whether or not reason truths also rigidly designate certain favoring relations across possible worlds, and so generate some kind of metaphysical necessity, is beside the point.

**B. Hume on the Is-Ought Gap**

The above argument bears some resemblance to Hume’s famous is-ought gap. Hume argued that you cannot derive an ought from an is:

> In every system of morality, which I have hitherto met with, I have always remark’d, that the author proceeds for some time in the ordinary ways of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when all of a sudden I am surpriz’d to find, that instead of the usual copulations of propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not. This change is imperceptible; but is however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given; for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. [3.1.1.27]

I reject one well-worn interpretation of this argument, according to which it is fallacious to reason from non-normative premises to normative conclusions. I think that one can gather

---

27 This is the reason conceivability arguments in the philosophy of mind have produced such controversy. One party argues that zombie worlds are conceivable, and based on that one then argues that phenomenal facts are non-physical facts. Another party replies that conceivability does not establish metaphysical possibility, for in some sense it is conceivable that water is not H₂O, and yet in another sense it is not metaphysically possible that water is not H₂O. This is a bit of a distraction. The most basic questions concern this world: is it conceivable that water is not H₂O given only the physical facts of this world and conceptual competence with respect to ‘water’? Is it conceivable that we are zombies given only the physical facts of this world and conceptual competence with respect to the phenomenal? It looks like the answer to the first is “no” and the second is “yes.”

28 Here I take it that ought facts are facts about what one has most reason to do. As a result, institutional ought facts that are not necessarily reason-implicating do not succumb to the argument. That is fine, for I only show that genuine normative reasons, and particularly the favoring relation constitutive of reasons, cannot be reduced to the natural-phenomenal world.
information about the natural-phenomena world—the is facts—and from those facts infer normative conclusions—the ought facts. When a utilitarian sees hoodlums burning a cat, for instance, he detects physical properties in the world. He might have an immediate non-inferential moral reaction to the situation, but he might also engage his theory to infer that burning the cat is wrong, that the hoodlums ought not do it. Similarly, when I consider the non-normative facts and the evidence for and against supernatural beings, I might infer that I ought not believe in God. These inferences are not fallacious. They are perfectly legitimate.

However, I do agree with Hume insofar as he claims that one does not deduce normative conclusions based only on the natural-phenomenal facts and normative conceptual competence. So what? What does that get us? With all the bells and whistles of the above argument, I think that the failure of deduction only establishes that the ought facts (and here Hume seemed to recognize that the normative facts are essentially relational) are not reducible to the is facts in the sense that ought facts require additional ontology to secure the truths of some normative statements. That, I believe, is the proper lesson of Hume’s famous is-ought gap.

I should note one other point of disagreement with Hume. He seemed to have a special skepticism for moral normativity, or perhaps practical normativity generally. The argument I have laid out, whether or not it leads to skepticism, applies to all reason and ought facts, not just practical, but theoretical reasons and oughts like those found in epistemology. So I cannot follow Hume in his practical skepticism, at least insofar as it is based on metaphysical concerns.

VII. Conclusion

We began with ethical naturalism as an attractive metaphysical thesis, but we have come to deny that normative reasons of any sort can be located in the natural-phenomenal world, and so a fortiori ethical reasons cannot be located in the natural world. Let me now come full circle
and say some words about how this connects up with varieties of ethical naturalism, and more generally how my conclusions impact other normative claims that are not explicitly about reasons.

It is important to return to ethics, for it is common for ethical naturalists to deny that agents have basic reason to be moral, underived from their subjective motivational states. That is, most ethical naturalists leave it open whether we have any reason to respond to their naturalized ethical properties. The first thing I want to note is that, in a very important sense, these ethical naturalists have not evaded the central point of this paper, for they still advocate reasons, if not underived ethical reasons, then practical reasons of some form or another, and if not practical reasons, then at least epistemic reasons. And the central point of this paper is that the favoring relation constitutive of all reasons, instrumental and epistemic included, cannot be naturalized. Ethical naturalism was thought to be attractive because it promised to tame an area of normativity that is uniquely problematic from a naturalist picture. But if ethical naturalists succeed, they uniquely immunize ethics from non-natural objections by sapping it of genuine normativity. Their other genuine normative commitments that are reason-implicating are not so fortunate, and they are the ones that run up against the attractions of a naturalized world-view.

Having said that, I do not think one should concede reasons internalism for practical reasons so quickly, for it seems that the very point and purpose of ethical theories of right action, good states of affairs, or virtuous character traits is to characterize those considerations that normatively relate to how individuals live their lives. Sincere ethical discourse carries with it normative commitments about which considerations provide reasons and which do not. The point of characterizing some acts as right, some states of affairs as good, and some character

29 Both Brink (1989) and Smith (1994) are kinds of reasons internalists.
traits as virtuous, for example, is to point out that these are acts that we have reason to perform, the states of affairs we have reason to bring about, and the character traits we have reason to develop. Otherwise, why spend a lifetime trying to distinguish these from wrong acts, bad states of affairs, and vicious character traits? Think of it this way: why should I care whether an act A is right or wrong unless these propositions are reason-implicating?

No doubt, with some effort we could use these ethical terms in entirely non-normative ways to pick out certain acts, affairs, characters, etc. without any presumption that these things are normatively significant, just as we now might use the word ‘taboo’ to pick out certain practices traditionally thought impermissible by Polynesian cultures, practices which no longer have any normative significance for us. But normal and important uses of ethical language—the ones we fight about—are normatively charged. Reasons are at the heart and soul of normative disciplines, and any putatively normative theory that has nothing to say about how we should think, act and feel is a shell of what we are after. In the normal, normatively engaged use of normative thought and language we claim, at least implicitly, that right acts are the ones to be performed, good states of affairs are the ones to be brought about, and virtuous characters are the ones to be cultivated. Those are the things we have reason to do, in contrast to the wrong, the bad and the vicious, which we have reason to avoid.

The same points can be made for epistemic discourse. The point of distinguishing those beliefs that are justified from those that are not is to distinguish beliefs that we have (epistemic) reason to form from beliefs that we have no (epistemic) reason to form (and perhaps epistemic reason not to form). As with much of ethical theory, much of epistemic theory is reason-implicating.
Naturalization projects nevertheless garner some rhetorical appeal through partial victory if and when they show that the relata of the favoring relations found in reasons can be naturalized. So, for instance, let us suppose that ‘good’ refers to pleasure only. Assuming pleasure is natural, so far so good. However, part of the point of calling pleasure good (as opposed to calling it bad) is to point out that it is the thing to be brought about, or pursued. Similarly, a reasons internalist who rejects any practical reasons that are unconnected to someone’s motivational set might naturalize one relata of his reasons, for the motivational states to be satisfied are no doubt natural.

Yet despite these partial victories the basic problem remains: the way in which these considerations normatively relate to us—the favoring relation—cannot be naturalized. A naturalized ontology, or a natural-phenomenal ontology for that matter, might include entities, properties, and relations wherefrom the favoring flows, and whereto the favoring points, but it will not include the favoring relation itself.

In closing let me simply recapitulate the argument I’ve defended in syllogistic form, and note some considerations for further thought.

1. If true reason claims are not deducible given all and only the actual truths of the natural-phenomenal world, along with conceptual competence of the normative, then MNR is false, absent some explanation for deductive failure that preserves MNR.
2. True reason claims are not deducible given all and only the actual truths of the natural-phenomenal world, along with conceptual competence of the normative.
3. There is no explanation for the deductive failure that preserves MNR.
4. Therefore, MNR is false.
5. If MNR is false, either no reason claim is true, or some reason claim is true, and our world contains a normative ontology above and beyond its natural-phenomenal ontology.
6. Therefore, either no reason claim is true, or some reason claim is true, and our world contains a normative ontology above and beyond its natural-phenomenal ontology.
We can also conclude that no reason-implicating discourse can be reduced to the natural-phenomenal world for that matter, and I have suggested that this would include ethical and epistemic discourse.

As with phenomenology, there is the option of eliminating any property that fails to fit into the natural world, and in some respects this is a more palatable option for reason relations than it is for phenomenal properties. Phenomenal properties are often thought to be manifest and undeniable, so eliminativism with respect to them might be an exercise in futile self-deception. However, favoring relations are not manifest in the way that phenomenology is; it might be easier to simply deny that there are reasons, realistically construed, than it is to deny that there is phenomenology. Here is the eliminativist motto: There are things that are and things that happen, but there is no fact about what things ought to happen (including no fact about what beliefs one should form regarding what does happen).

But reasons eliminativism as an ontological thesis is also costly. If there are no reasons, then it would seem that we have no reason to believe in the entities, properties and relations posited by ideal versions of the natural sciences, however psychologically compelled we might be. In general, we would have to consider how to be realists about some entities, properties or relations if we are irrealists about reasons for believing in those entities, properties or relations.

Alternatively, if we are realists about reasons, do we then need second order reasons for being realist about the first order reasons; wouldn’t we need evidence that reasons XYZ obtain rather than reasons ABC? If so, what is the metaphysical status of the second order reasons that provide evidence for the first order reasons, and how are we to avoid the looming infinite regress?
I have assumed that normative statements are truth-apt, and true in some suitably mind-independent way. As a last option one might develop an expressivism about reason-statements, including those found in epistemology, and one might even try their hand at an adequately truth-preservative expressivism, though I doubt this will respect our intuitions about mind-independence. Still, one would then have to consider how to be a realist about natural entities, properties and relations given an expressivist account of evidence for, and justified beliefs about, these natural things. I anticipate these to be vexed issues.
Works Cited


