Supervenience Arguments and Normative Non-naturalism*

Billy Dunaway
University of Oxford
forthcoming in *Philosophy and Phenomenolgical Research*

1 Defining non-naturalism

Frank Jackson (1998) gives an argument against familiar non-naturalist views about the normative which has been endorsed, in essentials, elsewhere in the metaethics literature. (The most prominent endorsements include Brown (2011) and Streumer (2008)). The primary aim of Jackson's argument is to establish the Descriptivism thesis, which is characterized as follows:

Descriptivism Every normative property is identical to a descriptive property.

Descriptive properties, for Jackson, are just properties which can be expressed with descriptive language—that is, with language that includes no normative vocabulary such as 'right', 'good', 'reason', etc.[1] Descriptivism thus implies that every normative property can be expressed using descriptive vocabulary only. (We will return in §2 to the question of *why*, according to Jackson et al., Descriptivism is supposed to be true.)

Most parties to the debate—both those friendly to Jackson's argument and those concerned to resist it—are willing to grant at the outset that *if* Descriptivism is true, then the traditional non-naturalist views about the normative found in Moore (1903) and elsewhere are false. Non-naturalism is true, in other words, only if Descriptivism is false.[2] Let us label this thesis Implication:

*Thanks to Campbell Brown, James Dreier, Allan Gibbard, Ishani Maitra, David Manley, Sarah Moss, David Plunkett, Peter Railton, Mark Schroeder, Bart Struemer, and Brian Weatherson for helpful discussion of the various issues covered in this paper.

[1] Jackson (1998, 113, 117). Gibbard (2003, 99) draws the same distinction using the term 'natural', stipulating that supernatural, mathematical and psychological properties count as "natural" in the relevant sense. Brown (2011) doesn't explicitly accept the same definition of 'descriptive property', presumably on the grounds that his version of the argument is supposed to avoid the "linguistic detour" present in Jackson's. He doesn't, however, offer an alternative characterization of the notion. I will not try to settle this question for Brown; but it should be clear that the points I make against Jackson's argument should apply *mutatis mutandis* to Brown's version *if* he were to accept the same characterization of what descriptive properties are.

[2] In addition to Brown (2011) and Streumer (2008), Shafer-Landau (2003, 94 ff.), Fitzpatrick (2008, 199), Shafer-Landau (2003), Suikkanen (2010), and Schmitt & Schroeder (2011, 146-7) all raise questions about a different premise in Jackson's argument. For more discussion, see §§5- 6 below.

IMPLICATION If DESCRIPTIVISM is true, then non-naturalism about the normative is false.

Jackson claims to find a commitment to the denial of DESCRIPTIVISM at the center of paradigmatic non-naturalist views. Speaking of Moore (1903), he says:

> What he really wants to insist on, I think, is an *inadequacy* claim: what is left of language after we cull the ethical terms is in principle inadequate to the task of ascribing the properties we ascribe using the ethical terms. He wants to object to exactly the claim I will be making.[3]

This, however, is not obvious given the descriptions non-naturalists provide for their own view. Moore, for instance, preferred (at one point) to explain his view in terms of the absence of a certain kind of *definition* of normative properties:

> When we say, as Webster says, 'The definition of horse is "A hoofed quadruped of the genus Equus," ' we may, in fact, mean three different things. (1) We may mean merely: 'When I say "horse," you are to understand that I am talking about a hoofed quadruped of the genus Equus.' [...] (2) We may mean, as Webster ought to mean: 'When most English people say "horse," they mean a hoofed quadruped of the genus Equus.' [...] But (3) we may, when we define horse, mean something much more important. We may mean that a certain object, which we all of us know, is comprised in a certain manner: that it has four legs, a head, a heart, a liver, etc., etc., all of them arranged in definite relations to one another. *It is in this sense that I deny good to be definable*. I say that it is not composed of any parts, which we can substitute for it in our minds when we are thinking of it.[4]

I do not wish to treat Moore's comments on the non-naturalist view as definitive.[5] What I do wish to point out is that there are plausible senses of 'definition', as Moore explains it, which are stronger than Jackson's inadequacy claim. That is: it makes sense to say that, even though descriptive language is adequate for expressing normative properties, it cannot provide a *definition* of those properties. And if we take Moore's talk of composition on board for a moment, it is easy to see why: even if descriptive language could pick out a normative property, it might not do so by delineating the *parts* of the property. Thus, given Moore's conception of definition, normative properties might not be definable in descriptive terms,

[3] Jackson (1998, 121)

[4] Moore (1903, 60), my italics.

[5] This is in part because Moore's own views on the topic were famously (and self-admittedly) confused. What he means by "definable" is not entirely clear (many will not find the language of composition helpful in a discussion of properties like goodness), and he retracted some his earlier claims about non-naturalism in his reply to C. D. Broad in Moore (1942). All I want to establish here that it is worthwhile and coherent to ask, in an investigation of Jackson's argument, whether IMPLICATION is true.

even though descriptive language is adequate for describing them.[6] The (limited) conclusion I wish to draw at present is simply this: some of Moore's claims suggest that the non-naturalist might allow that DESCRPTIVISM is true, and that IMPLICATION is, according to the non-naturalist, false.

In the subsequent sections of this paper I motivate and elaborate on this strategy for resisting Jackson's argument. I motivate the response by arguing that Jackson's premises cannot all be assumed in an argument against non-naturalism; someone who makes these assumptions without deducing them from a prior assumption that non-naturalism is false thereby incurs some implausible commitments. I then elaborate on one way of resolving these implausible commitments by rejecting IMPLICATION as a legitimate premise in an argument against non-naturalism. This is an approach that has gone unappreciated in the literature. The upshot is that a property-identity in the form of DESCRIPTIVISM might be something the non-naturalist can accept. This conclusion is, moreover, of relevance beyond localized debates in metaethics, and I close by briefly discussing its implications in any area of metaphysics where supervenience and reduction are at issue.

2 Jackson's supervenience argument

2.1 Rightness-entailing predicates

As mentioned in §1, Jackson argues against non-naturalism by arguing for DESCRIPTIVISM. His argument proceeds from the assumption that the normative supervenes on the descriptive, plus other apparently minimal "auxiliary" assumptions. I outline the argument below, before arguing on independent grounds in §§3-4 that these auxiliary assumptions are together not acceptable premises in an argument against non-naturalism.[7]

Jackson begins with the global supervenience of the normative on the descriptive:

GS $\forall w, w^*$: if w and w^* are exactly alike descriptively, then they are exactly alike normatively.[8]

[6] Analogies that are friendly to Moore abound here. Given the Ideal Gas Law, the volume of an ideal gas can be described entirely in terms of the amount, temperature, and pressure of the gas. But no one would suggest that volume of an ideal gas has pressure, among other things, as a constituent part. (Note that analogous reasoning would lead to the conclusion that pressure has volume as a constituent part.)

[7] Gibbard (2003, Ch. 5) also develops an argument along these lines, though the argument is adapted to a setting where the semantics for normative expressions is expressivist. And Kim (1978) originally outlined the form of argument, abstracting away from Jackson's concern with normative properties in particular. I return to more general issues in the concluding section of this paper.

[8] Jackson (1998, 119). For more discussion of various kinds of supervenience thesis and their relations to each other, see Bennett (2004).

He then argues that GS implies that normative predicates are equivalent to descriptive predicates. Here is his explanation of why this follows from a later paper:

> Consider any right action R_1. It must have some particular descriptive nature or other, as it is impossible to be right without having some descriptive nature or other. Let "x is D_1" is be the open sentence that ascribes that nature and also fully specifies descriptive nature elsewhere in R_1's world. It must then be the case that "x is D_1" entails "x is right" [...] Now consider any other right act R_2. With D_2 specified as for D_1 above but with "2" for "1", we get the result that "x is D_2" entails "x is right". From which it follows that "x is D_1 or D_2" entails "x is right". Repeating the process for every right act in logical space, we get "x is D_1 or D_2 or D_3 ..." entails "x is right". But, as we included every right act in logical space, the entailment must also run the other way. We have thus derived the logical equivalence of the infinite disjunctive open sentence "x is D_1 or D_2 or ..." with "x is right".[9]

That is: given a right action, there is a descriptive predicate D which describes the intrinsic features *and* worldly environment of that action. From GS, it follows that there is no other possible action which satisfies D and is not right. Call such a predicate *righness-entailing*. By disjoining each descriptive rightness-entailing predicate—one for each possible right action—we arrive at a big disjunctive descriptive predicate that is equivalent to 'right'.

The notions of entailment and equivalence, as I will use them here, are *modal*. For any predicates A and B, A *entails* B just in case for every possible world w, the set of objects B is true of at w includes the set A is true of at w; A and B are *equivalent* just in case these sets are the same at every possible world. Similar definitions of entailment and equivalence are available for properties: for any properties α and β, α *entails* β just in case for every possible world w, the set of objects that instantiate β at w includes the set that instantiate α at w; α and β are *equivalent* just in case these sets are the same at every possible world.

Extending this argument to show the equivalence of normative and descriptive properties, and not just predicates, requires a further assumption. For it might be that certain predicates fail to express a property (and perhaps the big disjunctive descriptive predicate is a candidate for such a predicate). So Jackson needs the following, which we can call *predicate-property correspondence*, or PPC:

PPC For every predicate P, there is a property α_P expressed by P, where for any world w, the set of objects P is true of at w is the same as the set of objects that instantiate α_P at w.[10]

[9] Jackson (2001, 655)

[10] This principle needs to be restricted to avoid paradox; I will assume that an appropriately restricted principle will license all of the uses to which PPC is put here and below.

2.2 Excursus: accidental expression

There are complications at this point, which Jackson and his followers recognize. Our §1 gloss on Descriptivism has it that, if P is a descriptive predicate, then the property α_P that P expresses (by PPC) is a descriptive property. But it won't do to say that *any* predicate containing descriptive language only counts as expressing a descriptive property. For instance, consider the predicate

D is the property I am actually thinking about now.

D is a descriptive predicate, as it contains no normative language. And by PPC, it expresses a property which, given what we have said, is a descriptive property. But this is problematic for the purposes of Jackson's argument. For suppose the property I am actually thinking about now is rightness; we then have an implausibly easy proof that rightness is a descriptive property. Something must be said to explain why this isn't sufficient to refute non-naturalism.

Jackson (1998, 119, fn. 10) and Streumer (2008, 538-9) restrict the notion of a descriptive property to those properties that are expressed by descriptive predicates that do not contain *property-denoting* terms of the form 'the property such that Φ'. That is, since **D** contains the expression 'the property I am actually thinking about now', the property it expresses fails to thereby count as descriptive. (Though it could, in principle, still be a descriptive property—so long as there is another descriptive predicate which expresses it without recourse to a property-denoting expression.)

This move is not only highly artificial; it is inadequate. For the same problems that arise for **D** also arise if we consider the predicate

D* is actually being thought about by me now.

D* contains only descriptive language, and in the appropriate circumstances, it expresses rightness. Intuitively, moreover, the explanations for why **D** and **D*** fail to express descriptive properties of the kind needed for Jackson's argument are the same. But **D*** contains no property-denoting expression. Another route is needed.

The natural thought to have here is that both **D** and **D*** do not count for Jackson's purposes because they contain indexical expressions like 'I' and 'actually'. We can then add a general requirement that descriptive properties be expressed *non-accidentally* by descriptive language. More precisely, let an expression e non-accidentally express the property P just in case, holding fixed the meaning of e, for every world considered as a context of utterance w_c, e in w_c expresses P. Thus on this approach, rightness may be expressed by **D** and **D***, but it isn't *non-accidentally* expressed, since there are worlds (considered as contexts of utterance) where **D**

and $\mathbf{D}^*$ fail to express rightness.[11]

With this amendment in place, the property that Jackson's big disjunctive predicate expresses still counts as a descriptive property. Since the rightness-entailing predicates from which the predicate is constructed contain no indexical-like terms, which property the big disjunctive predicate expresses is not accidental in the relevant sense. So at this point the argument can proceed as before, with our revised conception of a descriptive property in place.

2.3 Identity and Descriptivism

Jackson's argument, however, is for Descriptivism, which requires more than that rightness is *equivalent* to a descriptive property—rightness, according to the desired conclusion, must *be* a descriptive property. The road from equivalence to identity is easy for Jackson, who accepts the following thesis about the individuation of properties, which we can call the *equivalence thesis*, or ET:

ET For any properties α and β, if α and β are equivalent—i.e., if they share an extension at every possible world—then $\alpha = \beta$.

With these premises in place, Jackson's argument can be summarized as follows: GS guarantees the existence of descriptive a predicate equivalent to 'right'; PPC tells us that this predicate expresses a (descriptive) property, and ET implies that this property is the *same* property as rightness. Since we can repeat the same style of argument for any normative property, these premises imply Descriptivism. Given Implication, non-naturalism is false as well.

3 Descriptivism without supervenience

In this section I will argue that there is something wrong with this argument against non-naturalism. I will do this by showing that just by accepting Jackson's premises ET, PPC, and Implication, one is thereby committed to implausible claims. Showing that these consequences of this position are implausible comes in two steps. The first step is to show that, from the denial of GS—which we can

[11] It might be desirable to complicate the definition of accidental expression to require that P is accidentally expressed by e only if e on its actual meaning expresses different properties in different worlds *in virtue of the linguistic rules governing e*. The reason for this is that some theoretical terms e_t might express different properties in different worlds (while holding meaning fixed) on account of the fact that e_t expresses the property which best fits the theoretical role delineated by e_t. If different properties fit this role best in different worlds, e_t will, according to our first-pass definition, express a property accidentally. But e_t could be a term from a well-confirmed empirical science, and would in this case be a paradigmatic case of a term that expresses a descriptive property. The proposed revision, on which it is a necessary condition on accidental expression that a predicate expresses different properties in different worlds in virtue of the linguistic rules governing the expression, remedies this difficulty. e_t intuitively expresses different properties in different contexts in virtue of the distribution of the causal profiles of candidate satisfiers across modal space, and not in virtue of the linguistic rules governing e_t.

call ¬GS—the truth of non-naturalism can be derived. In schematic form, where NN is the non-naturalist thesis, this amounts to the following:

$$\frac{\neg GS}{NN}$$

Note that by claiming that non-naturalism is a consequence of ¬GS, I am not in any way claiming that ¬GS is plausible. The only claim in this part of the argument is that non-naturalism can be derived from an assumption of ¬GS.[12] Claims about what follows from a supposition in no way imply that the supposition is a plausible one.

The second step is to show that the someone who accepts ¬GS alongside Jackson's auxiliary would still be committed to the conclusion that non-naturalism is false, for reasons very similar to Jackson's original argument. That is, from ¬GS and the auxiliary premises, the falsity of non-naturalism can be derived; in schematic form, this amounts to the following (here I abbreviate Implication as 'I'):

$$\frac{\neg GS \quad PPC \quad ET \quad I}{\neg NN}$$

Jackson, of course, has already showed us how to derive ¬NN from GS and the auxiliary assumptions:

$$\frac{GS \quad PPC \quad ET \quad I}{\neg NN}$$

Thus deriving the denial of non-naturalism requires only assumptions of PPC, ET, and Implication. Someone who accepts the auxiliary premises alone is committed to the denial of non-naturalism.

These two steps together show that it is illegitimate to accept all of these auxiliary premises in an argument against non-naturalism. Given the second step, someone who accepts the auxiliary premises is committed to the denial of non-naturalism regardless of whether they accept GS. But the first step shows that one's commitments regarding the truth of non-naturalism should not be independent of whether one accepts GS, since if we were to accept ¬GS, we should be committed to non-naturalism and not its denial. Whatever claims one accepts to play the role of the auxiliary premises in an argument against non-naturalism should on their own be consistent with both non-naturalism and its denial; a plausible set of auxiliary premises will not by themselves permit a derivation of the denial of non-naturalism. But Jackson's premises have precisely this feature: if one were to accept them prior to deducing them from a denial of non-naturalism,

[12] By 'can be derived', I simply mean that it is an a priori consequence of the relevant assuptions. Thus by a schematic representation of an argument such as the one above, I mean to represent that the claim below then line is an a priori consequence of the claims above the line.

one would be committed to thinking that non-naturalism is false even if GS were false. Hence Jackson's PPC, ET, and IMPLICATION, even if true, are not legitimate premises in a supervenience argument against non-naturalism.

I explore diagnoses of why exactly this is so beginning in §5 after arguing in greater detail for each step of the argument outlined above.

3.1 Step 1: metaphysical consequences of failures of supervenience

Our first step is to argue that if GS is is false, then non-naturalism must be true. The aim is to show that the following holds:

$$\frac{\neg\text{GS}}{\text{NN}}$$

The negation of GS is the following claim:

¬GS $\exists w, w^*$: w and w^* are exactly alike descriptively, and there is some normative respect in which w differs from w^*.

¬GS requires, in other words, that the following obtain: there are two descriptively alike worlds which differ over whether some action is right—i.e., there is some world where an action is right, and a second world which is descriptively identical to the first, but where that same action is not right.

I am claiming that non-naturalism follows from ¬GS. This is because non-naturalism is supposed to be a view according to which the normative is, in some important sense, *independent* of the descriptive. We haven't explained what the appropriate sense of 'independent' is, but it is clear that whatever the appropriate sense is, the normative is independent of the descriptive in the relevant way if it fails to even *supervene* on the descriptive. One way to illustrate this is by drawing attention to similar debates in other domains. Take dualism about the mental, for instance: we might say that it is likewise an independence thesis, holding that certain aspects of the mental (perhaps *qualia*, the qualitative aspects of experience) are independent of the physical, biological, chemical etc. Chalmers (1996) defends this kind of thesis by arguing for the falsity of the following global supervenience thesis:

GS-MENTAL $\forall w, w^*$: if w and w^* are exactly alike in all physical, biological, chemical, etc. respects, then they are exactly alike mentally.

Hence much of the debate in this area centers around the metaphysical possibility of so-called "philosophical zombies": *if* they are in fact metaphysically possible, then GS-MENTAL fails and dualism is vindicated. Quite plausibly, the same is true

for GS and non-naturalism about the normative: *if* the supervenience thesis fails, non-naturalism is vindicated.[13]

While a denial of GS is much less plausible than a denial of GS-MENTAL, I only will be asking what can be derived from ¬GS if it is assumed as a premise. (Thus I will not be proposing that we argue for non-naturalism, and against Jackson, by arguing for ¬GS.) One can suppose things that one knows not to be true; for instance, one supposes the negations of logical truths when proving them by *reductio*. The arguments that appear below should be thought of along these lines, as they show what one would be committed to if one accepted all of the premises.

3.2 Step 2: descriptive rightness-entailing predicates under ¬GS

Our next step is to show that from the conjunction of Jackson's auxiliary premises and ¬GS, the falsity of non-naturalism can be derived.[14] The aim, in other words, is to show that the following holds:

$$\frac{\neg\text{GS} \quad \text{PPC} \quad \text{ET} \quad \text{I}}{\neg\text{NN}}$$

In outline form, the argument for this goes as follows: we can construct rightness-entailing predicates by using identity and reference worlds, and these predicates can be shown to be rightness-entailing even if we explicitly assume ¬GS. Once we have shown that descriptive rightness-entailing predicates can be constructed under ¬GS, the rest of Jackson's argument against non-naturalism can be repeated exactly as before.

Here, then, is the argument in greater detail. Letting $i_1, i_2 \ldots$ designate all possible instances of rightness, and $w_1, w_2 \ldots$ designate possible worlds in which there is at least one instance of rightness, there are rightness-entailing predicates of the following form:

[13]The terminology is tricky here, but this shouldn't obscure the underlying issues: Chalmers claims that his view, which denies GS-MENTAL, is nevertheless a version of *naturalism* (Chalmers (1996, xiii)). Thus it would be misleading to say that his view is an instance of non-naturalism about the mental. But this is because Chalmers has a distinctive theory about what 'naturalism' is, which may not capture the sense of 'naturalism' at issue in metaethics. I am assuming that there is still a clear sense in which the Chalmersian about the mental and the Moorean about the normative maintain that these domains are independent of others, regardless of the terminology we use to mark the similarity.

[14]We should note here that Jackson is surely right in saying that his preferred method for constructing descriptive rightness-entailing predicates requires GS. Suppose that ¬GS is true, and take the worlds w_1 and w_2 that differ only in normative respects. There is then some action which is the *same* with respect to its intrinsic descriptive features and environment, but which is right in w_1 but not w_2. So descriptive characterizations constructed according to Jackson's method will not be rightness-entailing if ¬GS is true. What this neglects is that there are other ways to construct the rightness-entailing predicates needed by Jackson's argument, and these don't require the supervenience assumption.

I $x = i_n$ and x is in w_j[15]

where i_n is an instance of rightness in w_j.

Predicates in the form of **I** are rightness-entailing, as no possible action satisfies the predicate yet fails to be right. This is so even if ¬GS is true. If ¬GS is true, then there is a pair of worlds—let them be w_1 and w_2—which differ only in whether some action i is right. Suppose that i is right in w_1; then, the predicate

$\mathbf{I}_1$ $x = i$ and x is in w_1

fails to be rightness-entailing only if there is some action that is not right, yet satisfies $\mathbf{I}_1$. But i in w_2 is *not* such an action—while i in w_2 is not right, it *also* fails to satisfy the predicate $\mathbf{I}_1$, as the non-right action is in w_2, which is distinct from w_1. Thus predicates in the form of **I** can be rightness-entailing even if there are pairs of worlds that differ only over whether a particular action is right.[16]

3.3 Excursus: refining supervenience

Let us suppose, for the moment, that rightness-entailing predicates in the form of **I** are also *descriptive* vocabulary. (I defend this assumption in §4.) There is an apparent tension in how we have described the situation: on the one hand, rightness-entailing predicates in the form of **I** are descriptive predicates. But, on the other hand, these predicates are supposed be able to pick out right actions even if supervenience fails, which is to say even if the only difference between the right action and a not-right action is a normative difference. We can't have it both ways; if predicates in the form of **I** are descriptive, then there is never only a normative difference between two actions. This might make it tempting to view our construction of alternative rightness-entailing predicates not as problematic for Jackson's argument, but rather as an indictment of our attempt to suppose that GS is false. The falsity of GS, we might conclude, turns out to be incoherent.

What this really shows is simply that more precision is needed in specifying the descriptive supervenience base for the normative. A GS-like thesis can play the exactly the same role in Jackson's argument, and can be coherently assumed to be false.

Here is a bit of terminology: there are some non-normative disciplines whose vocabulary is sufficient, in an intuitive sense, for giving a complete description of the supervenience bases for normative properties. To give a positive characterization of relevant supervenience bases, we will need, at the very least,

[15]Here I conflate predicates with open sentences. The open sentence can easily be converted into a predicate by the device of lambda-abstraction which is familiar from formal semantics. Where $\ulcorner F(x) \urcorner$ is an open sentence with the free (unbound) variable x, binding the free x with the lambda-operator, yielding $\ulcorner \lambda x.F(x) \urcorner$, denotes a function from objects to truth-values (namely a function which assigns 'true' to an object in case that object is F, and assigns 'false' otherwise). The lambda-abstracted expression therefore has the same semantic function as a predicate.

[16]I owe this point to Mark Schroeder, who initially made it in conversation, and also thank Campbell Brown for subsequent discussion of the issue.

microphysics. Further metaphysical investigation—which won't delay us here—might reveal that psychology, or perhaps theology, is needed too.[17] We can call the vocabulary from these disciplines *MPT vocabulary*, and can use it to formulate a minimal global supervenience base for normative properties. This gives us a revised version of global supervenience, namely:

GS_{MPT} $\forall w, w^*$: if w and w^* are exactly alike in all MPT respects, then they are exactly alike normatively.

By restricting attention to a global supervenience thesis formulated in terms of MPT vocabulary, we have a thesis whose denial is coherent. This is the thesis $\neg\text{GS}_{\text{MPT}}$, which is formulated as follows:

$\neg\text{GS}_{\text{MPT}}$ $\exists w, w^*$: w and w^* are exactly alike in all MPT-respects, and there is some normative respect in which w differs from w^*.

Not every way of combining pieces of MPT vocabulary yields a new piece of MPT vocabulary. For simplicity, suppose spacial predicates like 'to the left of' and logical expressions such as quantifiers and identity are parts of microphysical vocabulary alongside 'quark', 'spin', etc. Then the expression 'is to the left of a quark' is then composed out of microphysical vocabulary only, yet it need not pick out a microphysical thing. If the only thing to the left of a quark is a ghost, then our expression picks out a decidedly non-microphysical ghost. This is a case of a combination of microphysical terms not yielding a larger, complex microphysical term. In principle, the same can occur if we add psychological and theological terms into the mix. I elaborate on this possibility in the next section.

3.4 Clarifications and the way forward

It is worth summarizing what we have shown so far, this time using our refined supervenience thesis. The auxiliary assumptions Jackson uses in his supervenience argument are not together acceptable assumptions in an argument against non-naturalism. The first step in showing this is the argument that the truth of non-naturalism can be derived from $\neg\text{GS}_{\text{MPT}}$. The second step is the argument that from $\neg\text{GS}_{\text{MPT}}$ and the conjunction of Jackson's auxiliary assumptions, the falsity of non-naturalism can be derived. And the conclusion is that the conjunction of PPC, ET, and Implication is too strong to serve as prior assumptions in an argument against non-naturalism: the falsity of non-naturalism can be derived from these claims alone. Someone who accepts them as premises will be committed to

[17] If one thinks that psychology necessarily reduces to microphysics, or that the supernatural entities in theology are impossible, then one can omit these elements from our discussion in what follows, and focus on microphysics. One could also imagine scenarios where further additional vocabulary is required: for instance if higher-order biological or chemical properties are not reducible to microphysics, then one will need to supplement the list. For doubts about the possibility of expressing any substantive supervenience thesis about the normative, see Sturgeon (2009).

thinking that even if $\neg GS_{MPT}$ were true, the denial of non-naturalism would be true as well.

Some clarifications are in order at this point. The first is that IMPLICATION, as I am interpreting it, is a material conditional which is equivalent to the claim that it is not the case that both DESCRIPTIVISM and non-naturalism are true. This is the weakest premise that Jackson needs in order to validly conclude that non-naturalism is false given DESCRIPTIVISM. I do not wish to deny that one might additionally think that, if IMPLICATION is true, then it is necessarily true, or perhaps even is a trivial consequence of the non-naturalist view. I suspect that Jackson's characterization of the view as an "inadequacy thesis", which I noted in §1, supports the view that non-naturalism *just is* the view that DESCRIPTIVISM is false. But I will not assume this here, however, since stronger readings of IMPLICATION are unnecessary for the purposes of Jackson's argument. All he needs is that the material conditional IMPLICATION can, alongside PPC and ET, serve as premises in a good argument against non-naturalism. I have argued that these claims cannot together play this role.

This point about the interpretation of IMPLICATION leads to a second point about what the foregoing shows (and what it does not show). What it does show is that a package of auxiliary assumptions which includes IMPLICATION is too strong to derive the denial of non-naturalism, since this conclusion can be derived given the assumption of GS_{MPT} or its negation. What this does not show is that the conjunction of PPC, ET, and IMPLICATION is inconsistent and cannot coherently be accepted. On the contrary: suppose for the moment that non-naturalism can be known to be false for reasons independent of Jackson's argument. Someone who knows this might, moreover, accept DESCRIPTIVISM as a consequence of GS_{MPT}, PPC and ET. IMPLICATION is a trivial consequence of these views, and there is no incoherence in accepting it. All we have shown here is that the opponent of non-naturalism cannot deduce the falsity of non-naturalism on the basis of a set of auxiliary premises which includes IMPLICATION; this is what would commit her to the absurd conclusion that the denial of non-naturalism would hold whether or not GS_{MPT} were true. The opponent of non-naturalism who instead deduces IMPLICATION from DESCRIPTIVISM and the falsity of non-naturalism would not be in this absurd position. This is because she can concede that, given $\neg GS_{MPT}$ and the auxiliary premises, both non-naturalism and DESCRIPTIVISM can be derived. Hence she can conclude that the falsity of IMPLICATION would follow if $\neg GS_{MPT}$ were true. Thus there is a way of coherently accepting all of Jackson's premises, but crucially this requires deriving IMPLICATION from prior knowledge of the denial of non-naturalism.[18]

This is the core of my claim that Jackson's supervenience argument fails; all

[18]Schematically: for our naturalist who rejects non-naturalism for reasons independent of Jackson's argument, actual acceptance of the auxiliary premises including IMPLICATION induces no absurdity, because she accepts IMPLICATION only as a consequence of other claims including ¬NN:

$$\frac{GS \quad PPC \quad ET \quad \neg NN}{I}$$

that is left is to defend the claim that the rightness-entailing predicates constructed under the assumption of $\neg GS_{MPT}$ are descriptive predicates. I do this in the next section. Then, I turn to a diagnosis of why Jackson's auxiliary premises cannot be assumed in an argument against non-naturalism.

4 World-names and Ramsification

In the previous section I assumed that the rightness-entailing predicates in the form of **I**, when constructed under the $\neg GS_{MPT}$ supposition, are genuinely descriptive. These predicates are semantically capable of distinguishing between worlds that are identical in all MPT respects but differ normatively. Consequently, it would be very tempting to infer from this that these predicates items of normative and not descriptive vocabulary. But in this case such a temptation is misleading. (Readers who are not tempted to think this may skip to the next section.)

Here is a quick argument that predicates in the form of **I** are descriptive predicates, though not MPT predicates. Recall the complete list of microphysical, psychological and theological terms, or what we have been calling the *MPT vocabulary*.[19] Worlds as a whole can be described in MPT terms, and some of these descriptions are *complete*—i.e., every true claim about a world w with MPT vocabulary is made by a complete MPT description of w. Let D_w be such a description of w. We can then form a description which not only completely describes w in MPT terms, but also *says that* it is a complete MPT description of the world:

> x is D_w and x has no other MPT properties.

Call such a description *MPT-closed*.[20]

With $\neg GS_{MPT}$ in place, worlds that differ normatively will satisfy the same complete MPT-closed descriptions. For instance, the worlds w_1 and w_2, which differ only in whether some action i is right or not, satisfy the same MPT-closed description. w_1 and w_2, then, cannot be distinguished in MPT terms.

Since she should accept that the truth of non-naturalism is a consequence of ¬GS, she would also accept that the negation of IMPLICATION is a consequence of ¬GS:

¬GS	PPC	ET
NN	¬I	

This is the crucial point that someone who denies non-naturalism might accept IMPLICATION alongside the other auxiliary premises, but not in a way that makes them available as premises in an argument against non-naturalism. Special thanks to an anonymous referee for *Philosophy and Phenomenological Research* for pressing for clarification on the workings of this argument.

[19] Recall also that these are placeholders for whatever vocabulary is intuitively needed to formulate a minimal global supervenience base for the normative.

[20] If MPT descriptions like D_w are to give a genuine supervenience base for the normative, the MPT vocabulary must include the MPT-closure. What I am calling an MPT-closure is analogous to the notion of a physical description of a world with a "stop clause" found in Jackson (1998, 13). While not all negative existentials involving MPT vocabulary themselves constitute MPT vocabulary, I will assume that the closure clause is a special case.

What we can do, however, is supplement such an MPT-closed world-description to get a purely descriptive predicate that distinguishes w_1 from w_2. Begin with a "functional" characterization of rightness in other normative terms—i.e., a characterization of the *intra-normative* connections between rightness, having a reason, etc.[21] I don't want to take on any specific claims about what exactly the proper intra-normative connections of rightness are, but it is highly plausible that the something like following will be correct (those who disagree with the particulars can substitute their favored theory in what follows):

R1 Necessarily, $\forall x$, if x is right, then there is some reason to do x;

R2 Necessarily, $\forall x$, if x is right, and one can do x, then one has overall reason to feel guilt if one doesn't do x;

R3 Necessarily, $\forall x$, if x is right, then one has overall reason to blame someone who can do x, but does not.

Let a *Ramsification* of the intra-normative connections of rightness be the result of, first, replacing all of the normative terms in R1-R3 with distinct variables.[22] This leaves us with the descriptive open sentences R1*-R3*:

R1* Necessarily, $\forall x$, if x has F, then doing x has G;

R2* Necessarily, $\forall x$, if x has F, and one can do x, then one bears H to feeling guilt if one doesn't do x;

R3* Necessarily, $\forall x$, if x has F, then one bears H to blaming someone who can do x, but does not.[23]

Next, we conjoin the open sentences R1*-R3* and bind the variables that replaced normative terms besides 'right' in with existential quantifiers to obtain its Ramsification. This is the predicate **R**:

R $\exists G, \exists H$: necessarily, $\forall x$: if x has F, then doing x has G; if x has F, and one can do x, then one bears H to feeling guilt if one doesn't do x; $\forall x$, if x has F, then one bears H to blaming someone who can do x, but does not.

The Ramsification **R** is a descriptive predicate, as it is built up out of non-normative vocabulary only.[24] We can, moreover, use it to supplement MPT-closed

[21] See Ewing (1947, 148-9), Gibbard (1990, 51), and Scanlon (1998, 97) for theories in the spirit of R1-R3.

[22] See Lewis (1970).

[23] I have sightly modified the syntax of R1-R3 to accommodate F, G and H as first-order variables which take properties as assignments. This is for ease of expression; nothing essential hangs on the difference as we could easily have used second-order variables to accomplish the same task.

[24] Jackson agrees: see Jackson (1998, 141).

world-descriptions to obtain a descriptive predicate that designates rightness. Here is how.

Return to the simple case of the worlds w_1 and w_2, which are identical in all MPT respects and differ only in whether a certain action i is right. Both, then, satisfy a particular big MPT-closed description, which we can call $D_{1\&2}$. These worlds differ, but in a non-MPT respect, namely in whether i is right. What this amounts to is a difference in whether i in addition satisfies the Ramsified description **R**. w_1, where i is right, is a world which not only satisfies $D_{1\&2}$ but in addition is such that i has a further property that satisfies **R**. w_2, where i is not right, likewise satisfies $D_{1\&2}$ but is not such that i in addition has a further property that satisfies **R**. This amounts to a way of combining MPT vocabulary to get a descriptive specification of a non-MPT property.[25] It thus secures the conclusion that predicates in the form of **I** are not, under the assumption of $\neg GS_{MPT}$, normative vocabulary. The world-names can be replaced with purely descriptive vocabulary.[26]

At this point, it will be helpful to address a worry about the present way of proceeding. The starting point for this worry is the following observation: often, Ramsified descriptions suffer from the problem of being satisfied by *too many* properties. This observation, applied to the present issue, then suggests the following: our use of Ramsified descriptions of rightness will fail to distinguish worlds that differ only in whether a certain action is right, as these Ramsified descriptions are satisfied by other properties besides rightness. It is worth pausing to make two remarks about this worry.

First, the issue of whether **R** is satisfied by *non-normative* properties besides rightness is irrelevant here. This would be a problem if we attempted to designate rightness by straightforwardly expressing it with the predicate

R $\exists G, \exists H$: necessarily, $\forall x$: if x has F, then doing x has G; if x has F, and one can do x, then one bears H to feeling guilt if one doesn't do x; $\forall x$, if x has F, then one bears H to blaming someone who can do x, but does not.

But that is not what is going on here. Rather, the situation is the following: we are, in the first place, constructing MPT-closed world descriptions, and then afterwards using **R** to single out a further property, not mentioned in the MPT-closure, that

[25] *Cf.* our earlier example of a combination of microphysical vocabulary that specifies a decidedly non-microphysical, ghostly entity.

[26] To be precise, here are the proposed descriptive characterizations of w_1 and w_2:

> w_1 is such that
> (i) it is $D_{1\&2}$ and there are no other MPT properties that it instantiates;
> (ii) there is a further property Z such that Z satisfies **R** and i has Z.
>
> w_2 is such that
> (i) it is $D_{1\&2}$ and there are no other MPT properties that it instantiates;
> (ii*) there is no further property Z such that Z satisfies **R** and i has Z.

is (or is not) instantiated in the world in question. Once we have secured this way of descriptively specifying worlds, we then use **R** to specify rightness as a *extra* property (i.e., a non-MPT property) that actions may or may not instantiate. In short, non-normative MPT properties might satisfy **R**, but they don't satisfy the predicate 'x is a further, non-MPT property that satisfies **R**'.

A second version of this worry is more threatening to the present project, but is also more difficult to motivate. If there are other *normative* properties besides rightness that satisfy the Ramsification **R**, then our descriptive predicates will fail to distinguish worlds that are identical in all MPT respects but differ over whether a certain action is right. This would be significant problem for the claim that rightness can be specified descriptively even if the supervenience premise GS_{MPT} fails. But it is also not obvious that there are normative properties besides rightness that satisfy **R**.

One argument that Ramsified descriptions of normative properties are satisfied by multiple normative properties is found in Smith (1994, 48-56). The central descriptions of the color property redness, Smith says, will be satisfied by other color properties—just as red things cause red sensations, so yellow things cause yellow sensations, and so on. So Ramsifying the color-terms out from the description of redness will result in a description that is satisfied by redness but also yellow and other colors. He dubs this the "permutation problem". Smith then suggests that Ramsified descriptions of normative properties will suffer from an analogous permutation problem, but here we might demur. Rightness can be distinguished from many other normative properties non-normatively: for starters, it is a property of actions, whereas having a reason is the property of agents. There are, then, at least some non-normative descriptions of reasons and rightness that will survive Ramsification and allow us to distinguish these properties at the level of purely descriptive specification. While this doesn't constitute an argument that no version of the permutation problem can apply to normative properties, it does reduce the motivation for thinking that Smith-style permutations are guaranteed to appear in the normative domain. Those who wish to press the problem will need to find another motivating case besides that of color if they wish to claim that a Ramsified description of rightness can't do the job we need it to do here.

5 Does ET have to go?

Given $\neg GS_{MPT}$ as a premise, we can still construct a descriptive predicate expressing rightness, and implausibly deduce the falsity of non-naturalism. Jackson's argument has gone wrong by assuming each of PPC, ET, and Implication as premises. It is common in the literature on Jackson's argument to criticize ET; Fitzpatrick (2008, 199), Schmitt & Schroeder (2011, 146-7) Shafer-Landau (2003, 91), and Suikkanen (2010, 99-103) all take this route. I have no intention to defend Jackson's claim that necessarily co-extensive properties are identical against these

criticisms. Nevertheless, an ET-denying response is, by itself, unsatisfying as a response to Jackson. After developing this point in the present section, I develop a more satisfactory non-naturalist response that rejects the assumption of IMPLICATION in the next.

It is sometimes said, with Jackson's argument in mind, that non-naturalism *just is* the view that normative properties are not identical to natural properties. Shafer-Landau (2003, 91) says the "central metaphysical commitment" of non-naturalism is "a rejection of moral-descriptive property identities". This suggests a diagnosis of why Jackson's argument fails: it includes an assumption that properties cannot be fine-grained, which is exactly the kind of thesis the non-naturalist needs to formulate her view. Jackson's argument fails, on this way of proceeding, because in using ET it assumes the denial of the core metaphysical commitment of the non-naturalist view.

This would be a tidy diagnosis for the non-naturalist to give, but an understanding the core of non-naturalism to be a denial of normative-natural property-identities has been grossly unmotivated. What we would like to see, in motivating a rejection of ET, is a general view about the conditions under which necessarily co-extensive properties are distinct, from which it follows that a denial of property-identity is a substantial "core" metaphysical thesis. But non-naturalists have in general failed to provide such an account. The upshot will be that the ET-rejecting route is in principle workable but is not, when unsupplemented by further theory, a satisfactory diagnosis.[27]

Although a full discussion of the issue would be out of place here, consider by way of illustration one of the much-discussed putative counterexamples to ET. Shafer-Landau claims that triangularity and trilaterality are necessarily co-extensive yet distinct properties.[28] If this is true, then ET is false. We would have an instance where necessarily co-extensive properties are distinct. But this does nothing to suggest that a denial of ET is a part of the core metaphysical thesis of non-naturalism; the non-naturalist doesn't vindicate the core metaphysical commitments of her view *simply* by rejecting ET.

Non-naturalism is, among other things, a substantial (and controversial) thesis about the metaphysics of the normative. The view incurs substantial metaphysical costs—although if non-naturalists are correct these are costs we should be willing to pay since there are other benefits to her view of the normative that outweigh

[27] Discussion of non-naturalism in the literature hasn't entirely ignored this question. But the discussions do show (sometimes inadvertently) that it is a difficult task to complete in a way that satisfies the non-naturalist's needs. See Suikkanen (2010, 99-103), Fitzpatrick (2008, 199-200), Streumer (2008, 543-5).

[28] There are several arguments he gives here, beyond a simple appeal to intuition. One is that if trilaterality and triangularity are identical, then the properties of laterality and angularity should be identical too. But this isn't obvious at all: if laterality is the property of having at least one side, and angularity is the property of having at least one angle, then they are plausibly not even co-extensive with each other: a straight line could be said to have one side but no angles (*cf.* Jackson (1998, 127)).

these costs.[29] But this aspect of the metaphysics of non-naturalism isn't accounted for simply by rejecting the thesis we have been calling Descriptivism. One can see this by noting that rejecting the identity of triangularity and trilaterality doesn't have similar implications for the metaphysics of triangles. If one does reject the identity, one isn't taking on a theoretically costly, substantial metaphysical commitment about triangles—costs one should be willing to pay only if there are other theoretical benefits to the resulting view of triangles. Even if Shafer-Landau is right that triangularity and trilaterality are distinct, he surely isn't saying that a metaphysical thesis analogous to non-naturalism is true of triangles. (To be sure, one is taking on whatever theoretical costs come with positing additional properties by rejecting ET, but this isn't a substantial commitment concerning triangles specifically.) Analogously, then, by affirming only that normative properties are necessarily co-extensive with, but distinct from, descriptive properties one does not thereby affirm the central metaphysical commitment of non-naturalism.

It is important to be clear about what this does show, and what this doesn't show. What it does show is that the simple claim that Descriptivism (and hence ET) is false isn't enough to count as the "central metaphysical thesis" of non-naturalism. The non-naturalist needs a theory of when properties are, and when they are not, distinct. And it needs to be one that underwrites Descriptivism as a metaphysically substantial thesis about the normative. As the example of triangularity and trilaterality shows, our pre-theoretic conception of the conditions for property-identity is not one which guarantees that the falsity of Descriptivism is equivalent to a metaphysically substantial thesis akin to non-naturalism. This brings us to what the present discussion does not show: that the non-naturalist cannot supply the needed theory of properties. She very well could, but I won't explore all of the options here. Instead, without prejudging the prospects for such a project, I will explore in the next section a different way of diagnosing the failure in Jackson's argument. On this approach, the core metaphysical thesis of non-naturalism is consistent with Descriptivism, but entails that, according to non-naturalism, Implication is false.

The central motivation for this approach is our earlier argument that Jackson's auxiliary assumptions are not acceptable as premises in an argument against non-naturalism. I will sketch below an account on which Implication is false according to non-naturalism; this affords us a nice explanation of why both Descriptivism and non-naturalism *must* be true given $\neg GS_{MPT}$, and why Implication cannot be assumed alongside GS_{MPT}.

6 Diagnosis: non-naturalism as a fundamentality thesis

I will claim below that the two steps in the argument of §§3-4 together make it very natural for the to reject the use of Implication to derive the falsity of non-

[29] This way of characterizing non-naturalism is inconsistent with the claims of so-called "quietists" such as Scanlon (2003). I won't be addressing the viability (or coherence) of the quietistic brand of non-naturalism here; though for discussion see McPherson (2011).

naturalism. This will involve sketching a framework for the non-naturalist view which implies that IMPLICATION is false if non-naturalism is true. The starting point of this sketch is a popular development in recent (meta-)metaphysics. This is a notion of "metaphysical fundamentality" as developed in the metaphysics literature by David Lewis, Kit Fine, Jonathan Schaffer, and others. In principle these claims are separable: one might be convinced that the most satisfactory way to proceed for the non-naturalist is to reject IMPLICATION but decline to implement this approach in a fundamentality-centric way. I will not however explore the other options for filling out such a picture here.

6.1 'Descriptive' properties

A 'descriptive' property, as we have been using the term, is simply a property that can be expressed without using normative language. And according to IMPLICATION, if normative properties are descriptive in this sense, then non-naturalism is false. This terminology obscures some potentially important metaphysical distinctions. We argued in §§3-4 that, given $\neg GS_{MPT}$ as a premise, the identity of normative properties and descriptive properties can be derived with Jackson's other assumptions. Even if this is so, there might be more to be said about the metaphysics of the normative that can't be captured using the term 'descriptive property' and its cognates.

In particular: even if rightness is a descriptive property, the normative nonetheless has a kind of explanatory priority over the descriptive, under the supposition of $\neg GS_{MPT}$. The crucial point here emerges when we focus on *how* this descriptive predicate succeeds in expressing rightness. The natural explanation to give here is that the descriptive predicate succeeds in this setting because some actions are *right* and thereby have a further non-MPT property that satisfies **R**. The property of rightness, given $\neg GS_{MPT}$, appears to be a property that metaphysically explains facts expressed using the descriptive predicate **R**. Put simply, an action has a further non-MPT property that satisfies **R** *because* it is right, and not *vice versa*.

Simply lumping properties into those which are 'descriptive' and those which are not fails to capture this potentially interesting fact about $\neg GS_{MPT}$. If the explanatory profile of the normative as glossed above is what makes non-naturalism true under $\neg GS_{MPT}$, then we can say why DESCRIPTIVISM and non-naturalism are both true under this supposition: whether a property is descriptive or not is orthogonal to the question of whether it is a non-natural normative property. It is then possible that normative properties are descriptive, but that they play the kind of basic explanatory role which follows given a failure of supervenience. It is then a very natural hypothesis that that this basic explanatory role for the normative constitutes the core metaphysical commitment of non-naturalism. I expand on this suggestion, and its relation to ET and IMPLCATION, below.

6.2 *Metaphysical fundamentality*

What is the sense in which the normative can be shown to play a distinctive explanatory role given $\neg GS_{MPT}$? Here is one gloss on what it might amount to, which borrows from work inspired by the use of the notion of a "perfectly natural" property in Lewis (1983).[30] Subsequent developments of this idea have taken on both different terminology and substantive philosophical stances about the notion.[31] Most of what I say below, however, will focus on a few motivating ideas which run throughout these different developments, and which I will discuss under the heading 'fundamentality'. Fundamentality affords us an appealing way to implement the motivation for rejecting IMPLICATION sketched above.

The notion of metaphysical fundamentality is intimately tied to a distinctive kind of metaphysical explanation. Fine says,

> [T]he relationship of ground is a form of explanation; in providing the ground for a given proposition, one is explaining, in the most metaphysically satisfying manner, what it is that makes it true.[32]

Examples are bound to be controversial, but here is a fairly natural illustration of the idea: the ground for the presence of hydrogen atoms in some way consists in, or is metaphysically explained by, the subatomic. This suggests a picture on which the most fundamental includes what is picked out by terms like 'quark', 'spin', and the like. To give the "most metaphysically satisfying" explanation for the facts about hydrogen, one will need to use microphysical vocabulary of 'quark', 'spin', and the like.

Characterizing the fundamental in terms metaphysical explanations lends some intuitive support to the idea that the microphysical is among the most fundamental. But it doesn't necessarily imply that the microphysical *exhausts* the fundamental, and this is where our assessment of non-naturalism under $\neg GS_{MPT}$ comes in. Once we have the notion of the metaphysically fundamental under our belts, it is natural to say that it is a consequence of $\neg GS_{MPT}$ that some normative properties are metaphysically fundamental as well. Under $\neg GS_{MPT}$, rightness is among the fundamental properties needed for a most metaphysically satisfying explanation of all that is the case: we need rightness to give the requisite metaphysical explanation of some facts expressed using the descriptive predicate **R**.

This very naturally suggests an explanation of *why* non-naturalism can be derived from $\neg GS_{MPT}$: simply put, this is because the core commitment of non-naturalism is that the normative is metaphysically fundamental, and the fundamentality of rightness can be derived from $\neg GS_{MPT}$.

[30] See Fine (2001), Schaffer (2009) and Sider (2012) for developments of this idea under the headings 'Reality', 'ground', and 'Structure'. I will address some of the differences between these frameworks in §6.3.

[31] Also, there may be more than one notion in the area; see Barnes (2012) on this idea.

[32] Fine (2001, 22)

Importantly, on this view, it cannot be assumed that non-naturalism is false if Descriptivism is true; normative properties might be fundamental but nonetheless describable with non-fundamental descriptive predicates. We can prove that this is precisely the situation we are in given the assumption of $\neg GS_{MPT}$. Thus even given GS_{MPT} and Descriptivism, it still does not follow that the normative is not fundamental. Since Descriptivism is true regardless of whether the normative is fundamental—and hence the antecedent is Implication is guaranteed to be true—an assumption of Implication amounts to an assumption that its consequent is true as well, i.e., non-naturalism is false. Such a premise is false according to non-naturalism and quite clearly cannot serve as a premise in an argument against the view.

It is, however, legitimate to assume a thesis that is very closely related to Implication. If non-naturalism is just the view that the normative has the kind of explanatory priority described above—it is metaphysically fundamental, features in "most satisfying" metaphysical explanations, etc.—then we can safely assume that if the *most fundamental* characterization of rightness is in descriptive terms, then non-naturalism is false. This is the assumption that non-naturalism is false if the Fundamental Descriptivism thesis is true:

Fundamental Descriptivism Every normative property is most fundamentally a descriptive property.

So while Implication is not a legitimate assumption in a Jackson-style argument, the following thesis is:

Fundamental Implication If Fundamental Descriptivism is true, then non-naturalism about the normative is false.

Of course unlike Descriptivism, Fundamental Descriptivism is not a trivial claim to establish. This produces a simple explanation of why Jackson's original argument fails once we use only legitimate assumptions in an argument against-non-naturalism: while he succeeds in establishing Descriptivism, he is only entitled to the premise Fundamental Implication and not Implication. Since Fundamental Descriptivism doesn't follow from Descriptivism alone, the argument fails.

The foregoing is just a sketch of one way of to implement the preliminary observation of §6.1, according to which the distinction between descriptive and non-descriptive properties misses out on the metaphysically significant features of the explanatory profile of the normative. Since it is just a sketch, various versions of the idea might be implemented with different understandings of the broad notion of fundamentality, or without use of the fundamentality-centric framework at all. But it should be clear from what has been said here that Jackson's use of Implication is a very natural assumption for the non-naturalist to target.

6.3 A new solution?

Suppose we accept that Jackson's auxiliary premises are flawed because they ignore potential distinctions concerning the fundamentality of the normative. It might then be thought: this is just an instance of the ET-denying strategies glossed in §5. One might argue for this as follows: this diagnosis requires us to be able to say that normative properties such as rightness are more fundamental than the descriptive properties they are necessarily co-extensive with. And this straightforwardly requires that normative properties are distinct from the descriptive properties they are necessarily co-extensive with—after all, if the normative properties are fundamental and the descriptive properties are not, then there is a difference between them, and hence by Leibniz's Law they cannot be the same property. The fundamentality-based approach from the previous section is just an instance of the ET-rejecting responses that are so prominent in the literature.[33]

This characterization of the fundamentality-based approach to non-naturalism threatens to classify it merely as a variation on the existing approaches that reject ET. But it is misleading in several respects.

First, it is not true to say that this approach *requires* normative properties to be distinct from descriptive properties. This is only so if fundamentality is a property of properties—a coherent, but ultimately optional claim. To see why it is optional, we can take a page from the approach to fundamentality (or "Structure") in Sider (2012). On Sider's view—simplifying somewhat—the only facts are fundamental facts; there are no non-fundamental facts. However, *sentences* containing non-fundamental (non-Structural) expressions can nonetheless be true: they have truth-conditions at the fully fundamental level. On this picture, then, there is one set of truths—the fundamental truths—which can be talked about in multiple ways. (Sider calls this specification of truth-conditions a "metaphysical semantics".) One way is to use the canonical, "joint-cutting" language, which contains only expressions that match the Structure of reality; presumably one uses the expressions 'quark', 'spin', etc. to do this. But we can also use non-Structural expressions—expressions that do not carve at the joints like 'banana', 'mountain', etc.—to talk about these same fundamental facts.[34]

The Siderian position points to a way for the fundamentality-centric approach to avoid a rejection of ET. In this framework the non-naturalist can hold that normative expressions are among those that match the Structure of reality; there are fundamental normative facts. 'Right' appears alongside 'quark', 'spin', etc. in the catalogue of Structural expressions. These same fundamental normative facts can be spoken of using descriptive vocabulary as well. The difference is that, to use descriptive vocabulary to express these same truths one needn't use Structural, joint-carving expressions. Using a descriptive expression to pick out rightness is like using 'banana' to talk about what is really a bunch of microphysical

[33]Other ET-rejecting strategies in Suikkanen (2010) likewise appeal to Leibniz's Law.
[34]Sider (2012, 112 *ff.*)

goings-on. There is no need to reject ET on this picture; all of the facts that it countenances are facts containing fundamental properties. The distinctive non-naturalist component of this view is just the additional claim that rightness is among these fundamental properties and 'right' is a Structural expression.

This brings us to a second point. Even in a non-Siderian framework for fundamentality, where we must reject ET to account for distinctions in fundamentality, the fundamentality-based approach to non-naturalism still cannot proceed by rejecting ET alone. This way of proceeding needs to be accompanied by a rejection of Implication as well.

Suppose we grant that ET is false because some normative properties differ in fundamentality from their necessarily co-extensive descriptive counterparts. As we noted above, Implication is naturally accepted on the basis of the thought that non-naturalism just is the view that normative and descriptive properties are distinct. Thus, on this picture, non-naturalism is true if Descriptivism is false. But the falsity of ET and hence Descriptivism can implausibly be derived under almost any circumstances given these assumptions. Suppose, for instance, what amounts to the denial of the non-naturalist view on the fundamentality-based approach—*viz.*, that the normative is not fundamental, and there is some necessarily co-extensive descriptive property that is more fundamental than rightness. Then the denial of Descriptivism follows: under this assumption there is some difference in the fundamentality of normative and descriptive properties; hence the properties must be distinct. If we keep *Implication* on the grounds that non-naturalism just is the view thatDescriptivism is false, we will be committed to the truth of non-naturalism. But we made assumptions that would seem to be extremely friendly to its denial. The way to avoid this consequence in the present setting is to reject the thought that the core metaphysical commitment of non-naturalism is a denial of Descriptivism, and hence to give up on the motivation for Implication.

Thus if we accept a conception of fundamentality that implies ET is false, we *also* need a new conception of non-naturalism that is adapted to a setting where differences in fundamentality are relationships between distinct properties. Intuitively, once we allow for distinct necessarily co-extensive properties, the cases where non-naturalism is false are just those cases where not Descriptivism but rather Revised Fundamental Descriptivism is true:

Revised Fundamental Descriptivism Every normative property is necessarily co-extensive with a descriptive property that is more fundamental than it.

A full fundamentality-based response to Jackson's argument, then, cannot get by just by rejecting ET: it must also reject the assumption of Implication, holding instead that all Jackson is entitled to assume in his argument is Revised Fundamental Implication:

Revised Fundamental Implication If Revised Fundamental Descriptivism is true, then non-naturalism about the normative is false.

To summarize: we have seen that if we take the explanatory profile of the normative as the central commitment of non-naturalism, then rejecting ET is optional while rejecting Implication is not. The natural conclusion to draw if we are attracted to this line of thought is that the real implausibility in Jackson's auxiliary premises which makes them problematic lies with the assumption of Implication and not ET. This approach has the virtues of nicely explaining the first step of the §3 argument (that non-naturalism follows from a ¬GS) while blocking the argument in the second step (by explaining why Jackson is not entitled to the Implication assumption). And as an added virtue of this approach, we might see the resulting view as one way of implementing the Moorean idea with which we began in §1, according to which non-naturalism is committed to the impossibility of a certain kind of definition of normative properties. The normative, on this way of thinking, is indefinable because it is metaphysically fundamental.

7 Supervenience and reduction in general: a cautionary tale

I have sketched and motivated a response to Jackson's argument according to which non-naturalism is consistent with Descriptivism, the claim that normative properties are identical to descriptive properties. The problem with Jackson's argument, I have suggested, is that it assumes Implication and thereby leaves no room for descriptive but metaphysically fundamental normative properties. Remedying this defect by disallowing assumptions of Implication-like premises may have ramifications for metaphysical theorising about domains other than the normative. In closing, I will offer the briefest sketch of why this might be so.

Some arguments in metaphysics aim to establish that properties from different domains are identical. But if non-naturalism is, given some Jackson-like assumptions, consistent with the identity of normative and descriptive properties, the bare identity fails to have substantial metaphysical import for the metaphysics of the normative—it fails to even rule out the metaphysically extravagant non-naturalist view. This suggests that caution is needed when drawing metaphysical conclusions from claims about property-identity.[35]

[35]Here are some examples: Schmitt & Schroeder (2011, 145-6) observe that any supervenience thesis is committed to the existence of necessary connections between properties in the supervening set and subvening base. The best explanation for such necessary connections, they claim, holds that each supervening property is identical with a property in the subvening base. (See also Block & Stalnaker (1999, 24) and Kim (2008, 101).) And Kim (1989, 45) argues that once we derive the equivalence of properties from a supervenience claim, we cannot hold that properties in the supervening class are causally efficacious unless they are identical to subvening properties. The present point is simply that, even if these arguments are sound, more needs to be said if they are to yield metaphysically interesting conclusions.

For instance, one natural thought is that property-identities are metaphysically significant because a *reduction*, in a suitable metaphysical sense, immediately follows. Kim (2008) claims that this is not only true; it is obvious:

> There is no question about the reductive import of identity. If pain = N_1 [where N_1 is a specific neurophysiological state], there is no pain over and above N_1 [...] This is an open-and-shut affair if anything in philosophy ever is: Identities do reduce [...]
>
> [A]s far as reduction goes, nothing beats identities. That appropriate identities achieve reduction is intuitively obvious and beyond any philosophical second thoughts.[36]

But it is overwhelmingly intuitive that non-naturalism is a metaphysically interesting view in part because it *denies* that the normative is reducible. Insofar as non-naturalists can hold that normative properties are identical with descriptive properties yet not reducible, we need to be careful to check whether the background assumptions in place really do license this conclusion. Should they be analogous to those in Jackson's argument for Descriptivism, the present paper casts significant doubt on the Kim's claim that a reduction follows.

These closing suggestions are quite obviously not decisive—after all, Kim might have additional background assumptions in mind that license his inference from identity to reduction. One obvious respect in which this is so comes from our discussion in §5: if Kim might have an appropriate theory of properties in mind which does license inferring reduction from property-identity. (Of course it isn't obvious what this theory might be.[37]) What these considerations show is not that inferences from identity to reduction are invalid; only that there are lots of moving parts in play when the relationships between supervenience, identity, and reduction are at issue. The possibility of rejecting Jackson's Implication premise as outlined above shows that there are some classes of assumptions under which metaphysically interesting reductive theses do *not* follow from bare identity-claims. Whether similar issues arise in other areas of metaphysics is a question that deserves a second look.

References

Barnes, E. (2012), 'Emergence and Fundamentality', *Mind* **121**, 873–901.

[36] Kim (2008, 100; 113). Fodor agrees:

> Functionalists are required to deny that pain is *identical to* the disjunction of its realisers [...] And the reason they have to say *that* is that *otherwise multiple realization wouldn't be an argument against reduction*. (Fodor (1997, 155), Fodor's emphasis.)

[37] Alternatively, Kim and others might have an assumption about fundamentality in the background; perhaps they are presupposing that *if* mental states are identical to neurophysiological states, then these states are *most fundamentally* neurophysiological states.

Bennett, K. (2004), 'Global Supervenience and Dependence', *Philosophy and Phenomenological Research* **68**(3), 501–529.

Block, N. & Stalnaker, R. (1999), 'Conceptual Analysis, Dualism, and the Explanatory Gap', *Philosophical Review* **108**, 1–46.

Brown, C. (2011), A New and Improved Supervenience Argument for Ethical Descriptivism, *in* R. Shafer-Landau, ed., 'Oxford Studies in Metaethics, vol. 6', Oxford University Press, pp. 205–218.

Chalmers, D. (1996), *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.

Ewing, A. (1947), *The Definition of Good*, Routledge and Kegan Paul.

Fine, K. (2001), 'The Question of Realism', *Philosophers' Imprint* **1**(1), 1–30.

Fitzpatrick, W. (2008), Robust Ethical Realism, Non-naturalism, and Normativity, *in* R. Shafer-Landau, ed., 'Oxford Studies in Metaethics, vol. 3', Oxford University Press, pp. 159–206.

Fodor, J. (1997), 'Special Sciences: Still Autonomous after All These Years', *Philosophical Perspectives* **11**, 149–163.

Gibbard, A. (1990), *Wise Choices, Apt Feelings*, Harvard University Press.

Gibbard, A. (2003), *Thinking How to Live*, Harvard University Press.

Jackson, F. (1998), *From Metaphysics to Ethics*, Oxford University Press.

Jackson, F. (2001), 'Responses', *Philosophy and Phenomenological Research* **62**(3), 653–664.

Kim, J. (1978), 'Supervenience and Nomological Incommensurables', *American Philosophical Quarterly* **15**(2), 149–156.

Kim, J. (1989), 'The Myth of Nonreductive Materialism', *Proceedings and Addresses of the American Philosophical Association* **63**(3), 31–47.

Kim, J. (2008), Reduction and Reductive Explanation, *in* J. Hohwy & J. Kallestrup, eds, 'Being Reduced: New Essays on Reduction, Explanation, and Causation', Oxford University Press.

Lewis, D. (1970), 'How to Define Theoretical Terms', *Journal of Philosophy* **67**(13), 427–446.

Lewis, D. (1983), 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* **61**(4), 343–377.

McPherson, T. (2011), 'Against Quietist Normative Realism', *Philosophical Studies* **154**(2), 223–240.

Moore, G. (1903), *Principia Ethica*, 2nd edn, Cambridge Univsersity Press. All page numbers to the revised second edition, published in 1993.

Moore, G. (1942), A Reply to My Critics, *in* P. A. Schlipp, ed., 'The Philosophy of G. E. Moore', Open Court, pp. 535–677. All citations to the 3rd edition, published 1968.

Scanlon, T. (1998), *What We Owe to Each Other*, Belknap Press of Harvard University Press.

Scanlon, T. (2003), 'Metaphysics and Morals', *Proceedings and Addresses of the American Philosophical Society* **77**(2), 7–22.

Schaffer, J. (2009), On What Grounds What, *in* D. Chalmers, D. Manley & R. Wasserman, eds, 'Metametaphysics', Oxford University Press.

Schmitt, J. & Schroeder, M. (2011), 'Supervenience Arguments under Relaxed Assumptions', *Philosophical Studies* **155**, 133–160.

Shafer-Landau, R. (2003), *Moral Realism: A Defense*, Oxford University Press.

Sider, T. (2012), *Writing the Book of the World*, Oxford University Press.

Smith, M. (1994), *The Moral Problem*, Wiley-Blackwell.

Streumer, B. (2008), 'Are There Irreducibly Normative Properties?', *Australasian Journal of Philosophy* **86**(4), 537–561.

Sturgeon, N. (2009), Doubts about the Supervenience of the Evaluative, *in* R. Shafer-Landau, ed., 'Oxford Studies in Metaethics, vol. 4', Oxford University Press, pp. 53–90.

Suikkanen, J. (2010), Non-naturalism: the Jackson Challenge, *in* R. Shafer-Landau, ed., 'Oxford Studies in Metaethics, vol. 5', Oxford University Press, pp. 87–110.