Skepticism about Naturalizing Normativity: In Defense of Ethical Nonnaturalism

William J. FitzPatrick

Abstract: There is perhaps no more widely shared conviction in contemporary metaethics, even among those who hold otherwise divergent views, than that practical normativity must be capable of being naturalized (i.e., captured fully within a metaphysically naturalist worldview). My aim is to illuminate the central reasons for skepticism about this. While certain naturalizing projects are plausible for very limited purposes, it is unlikely that any can provide everything we might reasonably want from an account of goodness and badness, rightness and wrongness, and unqualified reasons for acting—at least if we are unwilling to accept certain deflationary or bullet-biting moves. Some naturalizing views can be shown to fail outright to capture the relevant normative facts or properties, while others have more promise but can also be seen to have certain limitations and costs, failing to capture elements that some of us take to be important to an adequate theory of practical normativity. There are, of course, far more naturalizing moves than can be considered here, so the aim is not to establish the truth of nonnaturalism through a process of elimination. But I hope to say enough to bring out the central worries about naturalizing projects and to pose some challenges that apply more widely, with the aim of showing that ethical nonnaturalism remains an attractive and well-motivated option at least for those of us who reject both nihilism and various forms of ethical deflation.

1 Introduction

There is perhaps no more widely shared conviction in contemporary metaethics, even among those who hold otherwise divergent views, than

that practical normativity must be capable of being naturalized. I understand this claim to mean that such things as goodness and badness, rightness and wrongness, and reasons for acting can all be captured entirely within a metaphysically naturalistic worldview—a conception of reality as containing only the sorts of entities and properties that are either susceptible to investigation by the empirical sciences or at least fully constructible from those that are. Such a claim thus stands in opposition both to the nihilistic denial of normativity altogether and to the nonnaturalist's insistence that normativity is real but can be captured only within a partly nonnaturalistic framework.

The most straightforward approach to naturalizing normativity would be some form of *normative naturalism*, which posits real normative properties and facts, takes them to be entirely natural ones, and holds that normative facts are represented by our true normative beliefs. Under this heading would fall everything from reductive identity views to non-reductive naturalisms, various forms of normative functionalism, naturalistic standard-based views, and so on. Certain constitutivist views will count as well. As Hille Paakkunainen (2014) has pointed out in recent work, a constitutivist might attempt to ground standards for good deliberation in constitutive principles of agency, and then understand reasons as (roughly) considerations that would be used as premises in pieces of good deliberation. This could provide a reductive naturalistic account of the property of being a reason—namely, as the property of *being fit to serve as a premise in a piece of good deliberation*, where this is cashed out as deliberation in accordance with constitutive principles of agency. Such a move would be analogous to the account of reasons offered by a teleological naturalist such as Philippa Foot (2001), though for her the standards of good deliberation come not from constitutive principles of agency but from natural teleology and facts about human needs and flourishing. Others might appeal to requirements of rationality understood in some non-constitutivist but still naturalized way: reasons would then be considerations that rationality requires us to treat as reasons in deliberation.[1]

The idea of naturalizing normativity in the broad sense I've defined would extend even further. Many *expressivist* views, for example, seem to count as well. On such an approach, normativity is again accepted rather than dismissed as illusory, and it is meant to be fully accounted for

---

[1] There is also a distinctive, radical version of constitutivism that isn't quite a form of normative naturalism but still falls under the broad naturalizing umbrella. On Christine Korsgaard's neo-Kantian approach, the point would not really be to identify normative properties and facts with natural ones, since on her view the whole model of discovering normative properties and applying normative truths in deliberation is a misconception of the practical sphere. Instead, normativity is construed in a radically practical way, ultimately reduced to practical necessity associated with agency—what one literally must do to function as a unified agent at all (Korsgaard 1996, 2009). Still, we might see this as a naturalizing of normativity, since normativity is embraced rather than denied, and it is accounted for in a naturalistic way, this time directly in terms of modal facts about agency.

naturalistically. Though Allan Gibbard (2003), for example, grants that normative concepts and thoughts are 'nonnaturalistic,' he means by this only that they have a directive, planning function rather than being used to represent the natural world as being certain ways, normatively speaking. Concepts themselves are natural things, and of course the only properties that exist for Gibbard (at least in a 'non-creeping-minimalist' way[2])—that is, the only features of the world—are natural ones. In fact, normative concepts have natural properties as their referents on his view—though for Gibbard (2006, 325–327) normative concepts aren't used to attribute those natural properties to things, as on a normative naturalist view. They are used instead to express plans, and that too is an entirely natural activity we engage in for reasons stemming from our natural, evolutionary history. So on this picture, normativity is embraced, despite the absence of normative properties and facts (except again in 'creeping minimalist' senses), and it is all accounted for naturalistically. That seems worth calling a 'naturalization of normativity' too, even if it is different from normative naturalism.[3]

There are other possibilities as well, but this is enough to illustrate the rich variety of naturalizing projects. To put my cards on the table, then: I am skeptical about all naturalizing projects in the practical sphere, at least where they claim to capture the most important forms of normativity, such as ethics and facts about unqualified reasons. I grant, of course, that there are limited forms of naturalization that work for limited purposes. For example, we can give plausible naturalistic accounts of normative claims that are just relative to stipulated or given ends and standards—as in the case of games, practices, roles, or functional objects. And we can give an attractive constitutivist account of a very weak form of the instrumental principle, which says just that we ought to avoid failing to take the known, necessary, available means to ends we continue to will. That normative principle can plausibly be grounded in the very nature of agency, for reasons Christine Korsgaard (1997) has brought out, involving the commitment taken on in willing an end. If willing an end essentially involves committing to take the necessary means, then in exercising agency at all we must in some sense be aiming to follow this weak, negative normative principle—or equivalently, the disjunctive positive principle that one ought either to take the necessary means or to cease willing the end. This seems to account for the 'ought' in question, which simply gives expression to that commitment and the related consistency requirement—again, all stemming just from

---

[2] On the idea of 'creeping minimalism,' see Dreier (2004), and the next footnote.

[3] If normative *concepts* and *thoughts* are granted, and normative thoughts can be called 'true' in a minimalist sense, and a true thought can be called a 'fact,' then we can speak equally of *normative facts*; and if the fact that murder is wrong implies that murder has the 'property' of being wrong, then there are *normative properties* after all, even for Gibbard. This is the 'creeping minimalist' sense in which Gibbard could give lip service to the normative realist's talk of normative properties, but it only muddies the debate.

the nature of agency. That is perhaps an interesting bit of naturalized normativity.[4]

Such plausible naturalizations of normativity are extremely limited, however. Even Mackie (1977), who advanced an error theory for ethics, allowed for true, naturalized normative claims understood as merely relative to given ends or standards or roles. And the weak instrumental principle supported by considerations about the nature of willing ends tells us nothing more than that we ought *either* to take the means *or* to drop the end, to avoid inconsistency with a commitment we've embraced. This in itself tells us nothing about what we have non-disjunctive, positive reason to do. It doesn't tell us, for example, that there is reason (even weak reason) for us to go ahead and take the means to whatever end we've willed, since it remains possible that all there is genuine normative reason for us to do in a given case is to drop the end (e.g., if it is evil to begin with). Indeed, not only does this sort of constitutivist approach yield only very weak results about instrumental normativity, but it also shows little promise of being capable of extension to ethical normativity more broadly, in any of the ways Korsgaard has attempted (FitzPatrick 2005, 2013).

Our real question, then, is whether there are naturalizing projects that hold genuine promise for giving us not merely accounts of some very limited forms of normativity, but everything we want in the way of practical normativity. That will of course depend on how much we want. If we are happy to give up various forms of normativity, or to settle for deflationary accounts of them, or to bite certain bullets about what reasons people have or lack, then naturalizing might well give us everything we want. But I confess to being greedy here, and the reason for this is straightforward: I am simply more convinced of the reality and robustness of various forms of normativity, such as categorical reasons or deeply non-relativistic moral requirements, and also more convinced of the falsity of the normative implications of certain naturalistic views, than I am of the truth of metaphysical naturalism. So I want a great deal from an account of normativity, and that is my starting point. This means that if naturalizing projects fail to deliver such results then I'm going to resist them, and I do resist them because I think they consistently fail to deliver everything we should expect from a proper account. Some naturalizing accounts have basic structural defects that cause them simply to miss normativity altogether. Others are more plausible and avoid a lot of the objections often raised against them, but I'll

---

[4] The normativity is naturalized insofar as the normative principle in question is fully accounted for in naturalistic terms, as a constitutive principle of agency—a principle one must in some sense at least be trying to follow in order to be exercising agency (e.g., willing an end) at all. If one wishes to speak in terms of what the relevant normative facts consist in, it might be said that the *fact that you ought either to do X or to cease willing E* consists in *the fact that in willing E you have committed to doing X, so that in failing either to do X or to cease willing E you will be involved in inconsistency*. Again, the claim would be that the 'ought' simply gives expression to this fact.

argue that they still go wrong in missing important aspects of normativity. Of course, I recognize that the worries I'll raise along these lines won't find much purchase with those who instead start out primarily committed to naturalism and are happy to follow it wherever it leads, even if that entails deflation or bullet biting. That is perhaps one of the unavoidable limitations of argument in this area, though it needn't impede all progress.

My ultimate aim here is to illuminate the central sources of skepticism about the prospects for naturalizing normativity, through a critical exploration of some representative naturalizing projects. Ideally, one might have liked to have a general, abstract argument that applies neatly to all naturalizing views and undermines them all at once, but I am not optimistic about the possibility of such an argument. Instead, I think we do better to proceed by examining some of the most prominent types of naturalizing strategy and exploring some revealing ways in which they seem to fall short. While this is not a fully general line of attack on the naturalization of normativity and cannot claim to constitute anything as strong as an argument from elimination in favor of nonnaturalism, it is also not merely a critique of the idiosyncrasies of a few selected views. Instead, the hope is to bring to light various difficulties that can be expected to plague a wide variety of naturalizing projects of the various types. If successful, this will at least pose serious challenges to be addressed by those who favor naturalizing normativity, and will provide at least tentative reasons for skepticism about whether these challenges can be met. It will thus also help to show that ethical nonnaturalism remains an attractive and well-motivated option for those of us who reject both nihilism and various forms of ethical deflation.

2 Avoiding a Common Misunderstanding of Nonnaturalism

It is worth starting by correcting a common misunderstanding of the motivation for nonnaturalism. Some naturalizers find it utterly mysterious why at least anyone who rejects nihilism would resist naturalizing normativity: positing nonnatural normative properties and facts strikes them as just pointless or even bizarre. That stance, however, is based on confusion and caricature of nonnaturalism. Frank Jackson (1998, 127), for example, imagines nonnaturalists thinking something like this: "I see that this action will kill many and save no one, but that is not enough to justify my not doing it; what really matters is that the action has an extra [nonnatural] property such that only ethical terms are suited to pick out."[5] But that is not at all what nonnaturalists think. We are not worried that the natural features cited as justification fall short as justifiers—as if the fact that an act will kill many and save none isn't enough by itself to justify refraining,

---

[5] Gibbard (2006, 328–330) gives virtually the same caricature of nonnaturalism in imagining how a nonnaturalist hedonist would add to the standard hedonistic justification for eating chocolate. (Cf. also Ridge 2014, 4.)

and requires supplementing with some "extra" justifying feature carrying the further credential of being nonnatural. That would indeed be silly (FitzPatrick 2008, 200; Scanlon 2014, 51). What we are questioning is instead something different: namely, whether the fact that the natural features do matter and do justify refraining from the action can itself be captured fully in a naturalistic way.

We grant that the natural features matter and justify refraining from the act, and so we aren't claiming that we need some extra property within the lineup of justifiers to do the real mattering and justifying—such as the imaginary nonnatural property Allan Gibbard (2003, 16), in a similar vein as Jackson, labels "exnat" and unsurprisingly finds peculiar. The suggested role of the nonnatural property or fact is very different: we're simply claiming that the fact that the natural features do matter and are wrong-making and do justify refraining from the action—*that* fact is not *itself* another natural fact. What is at issue here are facts about the normative significance of natural features of the world (Scanlon 2014, 21, 33)—or as Jonathan Dancy (2004) helpfully puts it, normative meta-facts like *the fact that* the fact that an action has these natural features makes it wrong or is a reason to refrain from doing it. Nonnaturalists are led to nonnaturalism because we don't believe such facts or normative meta-facts can be cashed out naturalistically—not because we hold that only special, spooky properties or facts really matter or justify anything (FitzPatrick 2008, 2011; Parfit 2011, 280).

So we may set aside the caricatures and focus on what is really at issue: *can facts about the normative significance of various natural aspects of the world be naturalized*? Or equivalently, we might put the issue in terms of properties, speaking of the normative meta-property, which some natural properties of things possess, of being right-making or wrong-making, or mattering, or justifying; and we can also speak of the resultant normative property of being right or being wrong, possessed by acts that possess the natural properties that have the normative meta-property of being right-making or wrong-making. Likewise, we can speak of normative properties of natural facts, such as the normative property certain facts possess of being reasons for doing something. The question, then, is again: can these various normative properties or meta-properties be naturalized? And if so, which approach is most plausible?

3 A Structural Error: Conflating Resultance Base Properties with Resultant Properties

Some naturalizing approaches go wrong from the start by just conflating what Dancy (1993, 2004, 2006) again helpfully calls *resultance base* properties with *resultant* properties, as by conflating good-*making* natural properties or facts with the property or fact of being good. This mistake

is evident in the familiar proposal that the property of being a good or right action *just is* the natural property of being happiness maximizing, for example, on the grounds that these normative concepts just name the same property we can also pick out with the non-normative concept 'happiness maximizing.' Some such thought is persistently tempting, especially to those who see the 'water'/'H_2O' model as a promising way to avoid Moorean open question arguments. But it involves a basic structural mistake that actually makes it incapable of capturing the evaluative or the normative at all.

I'll begin by exploring this structural error not because it characterizes all naturalizing views—it does not, as we'll see—but because exposing the error undermines a familiar range of naturalist views and helps us to see what a more adequate naturalism would have to try to do, which is a useful first step, setting up the critical questions that will be addressed in later sections. At the same time, discussion of these issues reveals the limitations of some prominent recent arguments against normative naturalism, which (as with the related arguments I offer in this section) are effective against naturalistic views that commit the structural error but not against more sophisticated views that do not. This is an important point to be clear about as well, so I will take some time to develop these points with a view to identifying the more promising forms of naturalism before moving on with the more general critique. I will then go on, beginning in section 7, to raise different problems for prominent representatives of those more structurally adequate views—problems that, while again illustrated using certain particular views, are suggestive of basic challenges faced by such views more generally.

The structural error I want to start with is easiest to see in a simple *non*-ethical context that all can agree to be a thoroughly naturalistic one—again underscoring the fact that the structural objections in this section are not against naturalism itself, but against certain tempting forms of it that even naturalists themselves should reject in favor of more structurally adequate views. Suppose you have a good computer. What *makes* it good are certain natural features it possesses *by virtue of which* it qualifies as a good computer. These are its good-making properties, such as fast processing speed, large memory, ability to handle several applications at once, etc. Call those natural, good-making features: *N*. Now it might be tempting to go on to say that the computer's being good just consists in its having the set of features *N*. But that would be a mistake. Though our concern here is not primarily with meanings, it helps to notice first of all that as far as the meaning of 'good' is concerned, there is little temptation here to speak of a special sense of 'good' where it just means: having fast processing speed, large memory, and so on—along with countless other special senses of 'good' when talking about cars or pianos or toothpicks. Far more plausible is the claim that 'good' has one sense in all these cases, meaning roughly: possessing features that make it satisfy the relevant standards of excellence

for things of this kind. That is, 'good' is a term that is used across a variety of contexts to indicate that the thing in question possesses features that play this role.

Again, our concern is with metaphysics, not meanings, but paying attention to how evaluative terms are actually used sheds light on what *evaluation* is, which in turn should guide our thinking about evaluative facts and properties. When we evaluate the computer as being good, we're not simply attributing the natural features N to it, as if calling it good were just another way of saying that it has N. What we're doing is more complex: we're saying that it has natural features that make it satisfy the standards of excellence for computers; and when we defend our evaluative judgment by citing those natural features N, we're citing them not simply as some set of features, but *as good-making properties* for a computer. We might explicitly say no more than: 'this is a good computer because it has N.' But that shouldn't mislead us into missing the complex implicit structure and supposing that all we're doing is saying that it has those features N. That would just miss the evaluation altogether. Someone could attribute N to a computer without having any idea these are good-making features, if he is ignorant of the standards for computers, and he obviously wouldn't be *evaluating* it. But if we tell him that it is a *good* computer *because* it has N, we're giving him new information: we're conveying that those features N are good-making for computers, making this one satisfy the relevant standards.

Likewise, two people can disagree on whether a computer is good even if they grant that it has N, and this isn't because they're disagreeing over whether it is N or some other set of natural features N_2 that is identical with the property of being good. What they're disagreeing about is instead whether N, as opposed to N_2, is really good-*making* in a computer; that is, they have a disagreement over the proper *standards*. Finally, notice that a property can be good-making in one context and not in another, where it might even be bad-making. Being sharp may be good-making in a knife and bad-making in a bookmark. So positively evaluating a knife is obviously more than just attributing sharpness to it. The *relation to relevant standards* is crucial.[6]

Even these simple cases therefore should make it clear that 'good' does not just function like a name picking out some set of properties like N, the good-making properties, as if evaluating something were simply attributing N to it. If 'good' is used to assert anything at all (as we're currently assuming) it is used instead to assert that something has good-making properties as such and thereby satisfies the relevant evaluative standards. And this in turn tells us about the evaluative facts and properties we're interested in. To spell out the above argument explicitly:

---

[6] The role played by standards on this model is similar to the role played by pure normative claims or facts on Scanlon's account (2014, 40–41).

(1) *Evaluating* some *X* as a good *F* is not merely to attribute certain features *N* to it, but to attribute *N* *as* good-making, thereby asserting not merely that *X* has *N*, but that *X* has *N* and thereby satisfies the standards of excellence for *F's*.

(2) If there are *evaluative facts* in such contexts, then they will be the facts that are asserted to obtain by corresponding *evaluations*.

(3) Therefore, the evaluative fact that *X* is a good *F*, if there is such a fact, will not simply be the fact that *X* has *N*, but the more complex fact *that X has N and thereby satisfies the standards of excellence for F's*.

Premise 1 is clearly supported by the above reflections on the simple artifact case. Premise 2 is just a conceptual point about what can plausibly be regarded as an 'evaluative fact': if an evaluation asserts that *p*, then surely if there is anything worth calling the corresponding 'evaluative fact' it is the fact that *p* (rather than some other fact that is *not* what is asserted by the evaluation as such). From this the conclusion follows: the evaluative fact, if it exists at all, is the complex fact—returning to our example— that the computer has features *N* and thereby satisfies the standards of excellence for computers.

Now there is nothing here that should inspire resistance from naturalists, since the complex fact I am identifying as the evaluative fact is still a fully *natural* one (assuming there is a naturalistic account of facts about standards of excellence for artifacts, appealing to facts about function grounded in designers' intentions, typical uses, and so on). All parties— naturalist or otherwise—should embrace the above model as a structurally more adequate model of evaluative facts. And this has straightforward implications for how we should think about the corresponding evaluative *properties*. The argument is again simple: If the mere fact that the computer has *N* is not itself an evaluative fact at all, then neither is the corresponding property of having *N* an evaluative property at all. As before, though, there is a related, more complex property that does have an obvious claim to be considered the evaluative property, namely: the complex property of *having N and thereby satisfying the relevant standards* (i.e., the property corresponding to the evaluative fact identified in the argument above). To take another example, a knife's being good obviously doesn't consist simply in its being sharp: otherwise, the hazardous bookmark would be good too. The evaluative property we are trying to capture is the complex property involving relations to appropriate standards.

Again, there is no reason stemming from naturalism to resist this: we are so far just pointing out the *appropriate* naturalistic identification of the evaluative property, which follows from the earlier identification of the corresponding evaluative fact. Perhaps some, however, will be inclined to resist this point, holding out for the simpler property identification I've rejected. They might claim that while *evaluation* indeed involves the

complex judgment I've described, and the *evaluative fact* is the complex fact proposed, and the *concept* of being a good *F* has the same complex structure, nonetheless the evaluative *property* itself should just be taken to be the simple property of having N.[7] But for the reasons given above, it's hard to see any good motivation for choosing that simple property *over* the complex one *if* both are recognized as legitimate candidate properties: the complex one clearly fits better with evaluation and with the evaluative fact and concept, as we've seen, whereas the simple one (the mere property of having N) is divorced from them and ostensibly has no more claim to being an *evaluative* property itself than the simple fact (that *X* has N) had to being an evaluative fact.

Of course, if someone had independent metaphysical reasons to be suspicious of complex properties of the sort I've proposed, while accepting simple properties, then she would have reason to reject my positing of evaluative properties identified with the complex properties in question. But in that case she should then simply reject the existence of evaluative properties altogether, rather than opting for just misidentifying them with the simple properties in question. The very plausibility of evaluative properties (at least in simple cases like this), however, is itself a very good reason for rejecting such metaphysical scruples.[8] Again, the best position for naturalists to take here is to abandon the overly simplistic identification—which just confuses the *good-making* property with the property of *being good*, leaving us with a property (having N) that has no plausible claim to being an evaluative property at all—and accept the more complex one. The real question when it comes to *ethical* evaluation, then, will be whether there is a satisfactory parallel in the ethical case to the more complex naturalistic identifications. That is the deeper challenge (to be taken up starting in section 7).

4 A Related Misidentification: Massively Disjunctive Natural Properties

Notice that the above points against what I've called the *structural error* apply equally against views that would identify goodness with some Jacksonian, massively disjunctive property *D*, where each disjunct is some subset

---

[7] Jamie Dreier has pressed this line in conversation, on behalf of those who wish to make the simple naturalistic identification I am attacking. My response in the remainder of this section is further developed in connection with the critiques of Jackson's, Gibbard's, and Smith's naturalist moves in sections 4–6 below.

[8] In the next section, I consider a different metaphysical inspiration for identifying evaluative properties with resultance base properties, which appeals not to the rejection of complex properties of the sort I've identified but instead to a *sparse conception of properties* that disallows the evaluative property's being anything over and above a disjunctive property constructed from resultance base properties. I take my arguments here to help to show why such a conception of properties is misguided, bringing to light another counterintuitive implication of such a view.

of properties such as *N* possessed by a possible good individual of the relevant kind (cf. Jackson 1998). While it is true that all and only good computers, for example, would possess *D*, *evaluating* a computer as a good one is no more simply attributing *D* to it than it was attributing *N* to it, and again this has implications for how the evaluative facts and properties should be identified. The trick of manufacturing a natural property that is coextensive with an evaluative one doesn't show that being a good computer is the same property as having *D*, for exactly the same reasons brought out in the previous section. Such moves again just miss the evaluative altogether by missing the implicit structure of the evaluative properties and facts and conflating *good-making* properties with the property of *being good*.[9]

Again, defenders of such views can always appeal to independent metaphysical principles to try to motivate them. In this case (or at least in the parallel ethical case) they will appeal to a 'sparse conception of properties,' and claim that *necessarily coextensive* properties are *identical*, thus motivating the claim that the goodness of an *F just is* the massively disjunctive property *D* with which it is necessarily coextensive, rather than being anything distinct from it. But in addition to the arguments others have given against such metaphysical claims, the arguments I have been giving here against such property identifications when it comes to evaluative properties may be seen precisely as showing why such metaphysical claims are dubious in the first place: if the above observations are intuitively compelling, as I believe they are, then clearly even necessary coextension doesn't entail identity, since the resultant evaluative properties (such as being a good *F*) plainly remain distinct from the resultance base properties (such as having *N* or *D*).[10]

Gibbard, too, makes a similar error, using an agent's "hyperplan" for what to do in all possible circumstances to concoct a massively disjunctive property *D* that an act can have—namely, the property of being an act of type A_1 in circumstances C_1, or A_2 in C_2, and so on. He then identifies this disjunctive property *D* with the property of *being what one ought to do* (Gibbard 2006, 324–327). That is just the same mistake again, in this case using the idea of a hyperplan to try to capture with at least some plausibility the *normativity* in the idea of being what one ought to do, but then *dropping* the hyperplan from the picture when it comes time to identify the property itself, and focusing just on *D*, the natural disjunctive property associated with the hyperplan. There is, however, no plausibility in the idea that *D* itself, which has nothing of the *normative* about it, *is* the property of being what one ought to do. Any hope of capturing the latter property must capture its normative nature, which *D* manifestly fails to do.

---

[9] For a similar critique of Jackson, see Dancy 2004. I have also discussed this in FitzPatrick 2008, 2011.

[10] Among other recent critiques of Jackson's claim that necessarily coextensive properties are identical, see Plantinga 2010 and Parfit 2011, 296–297.

What Gibbard should claim is identical to being what one ought to do (given the rest of his view) is not *D* itself, but rather the complex property of *possessing D and thereby being consistent with a certain hyperplan*. That would bring whatever plausibility the appeal to hyperplans has for capturing normativity back into the picture, so that this more complex property is at least of the right *sort* to be a candidate for the property of being what one ought to do (though again for Gibbard normative judgments aren't a matter of attributing that property to acts, which would be normative naturalism, but are instead a matter of expressing plans). Notice, however, that this more complex property is clearly *plan-relative*, which means that such an account hardly succeeds in capturing the trappings of realism, as advertised. Realists (whether naturalist or nonnaturalist) aren't going to be very excited about an account of the property of being what one ought to do that is merely relativized to an agent's hyperplan (FitzPatrick 2011). So Gibbard's expressivist attempt to mimic a plausible naturalist realism just commits the same sort of structural error I have identified in simpler cases, and if it is modified to avoid that error by incorporating the relevant structure within his framework then it makes the property of being what one ought to do merely plan-relative.[11]

5 The Limitations of Parfitian Objections to Naturalism

My critique of the structural error so far dovetails nicely with Derek Parfit's Triviality and Normativity Objections (Parfit 2011). Parfit might point out that if a computer's being good just were its possessing *N*, then it would be trivial to say that this *N* computer is good, in the sense that we would be adding no new, non-linguistic information with this evaluative predicate. But of course it's *not* trivial to say that this *N* computer is good. Similarly, he might note that the simple identification (of goodness with N-ness) would preclude our being able to say that the possession of *N* *makes* this computer have the distinct property of being good. But we *can* say that the possession of *N* makes this computer have the distinct property of being good. And *evaluating* a computer as good is clearly a different thing from saying that it has *N*, which again is hard to explain on the simple identification view. So Parfit's points reinforce my own.

At the same time, however, focusing on a clearly naturalistic example like the computer reveals the limits of these structural objections. What they show is just that if one is going to naturalize normative properties, one has to avoid the sorts of misidentifications I have flagged, such as conflating the property of being good with the good-making properties.

---

[11] Even apart from the point in the text, there are deep worries about a thoroughgoing kind of relativism plaguing views such as Gibbard's, despite expressivists' familiar attempts to evade them. These worries remain even if my critique of the structural error is rejected. I develop this line of objection to expressivism in FitzPatrick 2011.

This doesn't show that the normative properties cannot be naturalized at all. Indeed, in the case of the computer they plausibly can be, and I've offered the standard-based account as a way to do it. The standard-based model, focused on the more complex property we have identified, clearly avoids the Parfitian-type worries. On the standard-based account, it is (appropriately) not trivial to say that this N computer is good, since we are saying (non-trivially) that by virtue of possessing N this computer satisfies the standards of excellence for computers (not merely that this N computer is N); and we do hold that the computer's being good is a distinct property from its having N, where N makes it have the property of being good, and we recognize that evaluating the computer is different from just attributing N to it. Yet the standard-based account accomplishes all of this and avoids the Parfitian-type worries, in this sort of case at least, within a plainly *naturalistic* framework.[12]

Similarly, then, when we move to the ethical cases that are our real concern, my structural objections and Parfit's arguments likewise tell against certain misidentifications, but they do not yet tell against *naturalism* itself. Call the property of being happiness maximizing '$N*$.' We should indeed reject the idea that calling an act 'good' or 'right' is just attributing $N*$ to it: that would again just miss the element of evaluation altogether. Even if $N*$ is the good- or right-*making* property of acts, it is not the property of being a good or right act, any more than N was the property of being a good computer. On the standard-based model I have proposed, the property of being a good or right action would instead be the complex property of having natural features that make it satisfy the relevant standards of excellence for human action. Whether this complex property is itself *natural* or not remains so far a wide-open question. It can't simply be assumed to be natural just because $N*$ is, but it is not automatically precluded from being natural either. Everything depends on whether the relevant standards can be fully naturalized in the case of ethics. If they can, then the worries so far are answered.

The task for the naturalizer is therefore to show, just as we would do in the case of artifacts, that there is some adequate, purely naturalistic account of appropriate standards of excellence for human action (or equivalently, of the good-*makingness* of $N*$). That is, she really needs to show three things, if this is to address our concern with ethics:

---

[12] I don't mean to imply that Parfit actually has worries about naturalizing normativity in the case of artifacts. The point is just that the *kinds* of worries he raises against naturalism in the cases he is interested in can equally be raised against certain forms of naturalism even in the case of artifacts, showing them to be inadequate. But since there are other, more adequate forms of naturalism in the case of artifacts (such as the standard-based account), this shows that the kinds of worries he raises don't automatically rule out *naturalism* as such in the cases he is interested in, but only the versions that commit such structural errors. Further argument is needed to raise problems for the more complex and plausible forms of naturalism, as I go on discuss later.

(i) that there are naturalistically derivable standards for human conduct,

(ii) that these are the standards we have reason to take seriously as appropriately governing our lives, as ethical standards are supposed to do in a very deep way, and

(iii) that this normative fact about them—their authoritative status as standards appropriately governing autonomous agents like ourselves—is itself just another purely natural fact (or alternatively, can be given an expressivist treatment without undue deflation).

This is where the real work for the naturalizer lies—not merely in telling some semantic story about how the reference of 'good' or 'right' is fixed, as if these terms worked anything like 'water' or 'heat.' And while the debate over the success of such a project may be a metaethical one, it isn't as abstract and removed from ethics as a lot of current work in metaethics is, because it will take substantive *ethical* reflection to determine whether a given naturalistic proposal for grounding the appropriate standards for human action is ultimately satisfactory or not. It is not enough in doing metaethics just to do semantics, metaphysics, and epistemology as semanticists, metaphysicians, and epistemologists who happen to be turning their attention to the ethical domain. The inquiry has to be ethically engaged—a point I'll come back to.

6 Functionalist or Explanatory Role Views

Before turning to my reasons for skepticism about naturalizing ethical standards, it is worth looking at another move that might seem to allow for the simpler naturalistic identification after all. I have in mind here functionalist or explanatory role approaches, where a normative concept such as rightness is said to refer to the natural property of acts, whatever it is, that plays some special role. There are a variety of possibilities here depending on the role specified in the reference-fixing description. The role could be as simple as that of causally explaining our use of the word 'right,' or it could involve causally explaining certain sociological effects of actions, or causing us to have certain psychological responses under certain hypothetical conditions. Michael Smith (2004, 196–199) has argued that such explanatory role views really combine an *a priori* element with an *a posteriori* one: they rely on an *a priori* claim about the connection between the concept and the role figuring into the reference fixing description; then it is an *a posteriori* matter to discover which property in the world fills that role.

Any such view will of course have to make it plausible that a natural property's filling the role in question qualifies that property to count as *rightness*, capturing its normativity. But if that challenge can be met, then naturalists might claim that talk of rightness was all along just referring to

whatever natural property N^* plays the relevant role. This would be like saying that talk of redness was all along referring to whatever microphysical property of objects plays the role of causing them to look red to normal perceivers under standard conditions (Smith 2004, 197). The addition of this functionalist structure would then avoid at least some of the earlier worries with views that just appeal to co-reference, and apparently it would do so without bringing in standards. For example, the functionalist account can explain how a judgment of rightness is indeed relevantly different from mere attribution of N^*, with the normative judgment adding information about the *role N^* plays*, just as on the standard-based model we are adding information about relations to standards. This might therefore seem to save the identification of rightness with N^* after all, so that being right *just is* being happiness maximizing, say.

It is doubtful, however, that such an identification is really supported by this sort of move.[13] Why, after all, is N^* relevant on this model? Not simply in itself, as maximizing happiness, but because it plays the special role in question: for example, on Smith's view it plays the role of being the property "we would desire our acts to possess if our desires formed a set that is maximally [empirically] informed, coherent and unified" (Smith 2004, 202). That is the lens that focuses the spotlight on happiness maximizing in the first place, and that lens was crucial to the concept of *rightness*. So if the project is to identify the property that constitutes rightness itself, the natural move here—parallel to what was earlier argued to be the appropriate move to make in identifying the evaluative property in an artifact, or in identifying the property of *being what one ought to do* within Gibbard's scheme—would be to incorporate this role into the proposed property. That would make the property a plausible candidate for actually *being* the property of rightness rather than merely being a property *relevant* to rightness (as N^* is). That is, the natural move here would be to identify the property of being right with the complex property of *being N^* and thereby being what we'd desire our acts to be like if our desires were suitably 'rational'*. By building the relation to rational desires into the property we at least put forward a property that has some claim (though ultimately not enough, I'll argue below) to capture the *normativity* of being right. And, as before, this more complex property remains as naturalistic as N^* is, so there are no reasons stemming from naturalism to resist this more complex identification.

The question, then, is why someone taking Smith's functionalist approach should reject the above move in favor of his narrower identification of rightness with N^* itself. In light of what has been said earlier, it is hard to see any good motivation for this, and the reasons against it are clear—indeed, this case is directly analogous to Gibbard's move and is

---

[13] Dancy (2006, 143) raises a worry about Smith's identification that is very similar in spirit to the objections in what follows.

objectionable on the same grounds. The only reason Smith can plausibly claim to be investigating the *normative* property of rightness at all—rather than just talking about familiar non-normative natural properties—is that he has introduced the apparatus of a role involving rational desires. That apparatus may then be used (analogous to Gibbard's use of the apparatus of a hyperplan) to identify a natural property such as N^* that is *relevant* to rightness (by playing the role bound up with rightness). But if we then drop that apparatus and focus simply on N^* as such (i.e., simply the property of maximizing happiness itself), which the apparatus pointed us to, we are thereby dropping the (allegedly) normative element from our focus, and the property we are championing is not the normative property we were after at all (i.e., it is not rightness, but just a different property that is relevant to but distinct from rightness).

Again, what we're really after in seeking rightness within a functionalist framework is not N^* as such, but N^* *as playing the relevant role associated with rational desire*, which leads instead to the more complex property identification I have proposed on Smith's behalf as a better form for his naturalism to take. But now notice that when revised this way the view then just becomes a version of the standard-based view, offering an *idealized subjectivist naturalization* of the standards: an act's being right consists in its according with the standards of proper human conduct, which it does by being happiness maximizing (the right-making property), where the standards spring from rational desires. So we're back, in effect, to the standards model, the real question then being again whether naturalism gives us the resources to capture normativity on that model without undue deflation or bullet biting.

Some might try to save the simpler identification by again driving a wedge between evaluation and the nature of the evaluative properties. That is, they might grant that there is a difference between calling an act right and saying that it is N^*: only the first is *evaluating* it, by saying or implying that the act has a property that plays a certain role, thus going beyond merely saying that it is N^*. But they will claim that the evaluative property itself is simply N^*, not the more complex, structured property I've suggested. As discussed earlier, however, it's implausible to suppose that we can capture an *evaluative* property even while stripping away precisely the elements purportedly distinctive of evaluation and evaluative facts and evaluative concepts—such as relations to relevant standards, relations to hyperplans, or relations to rational desires. When these elements are left out of the property, so that it is stripped down merely to something like N^*, what remains is just a plainly *non*-evaluative property, and we have then simply missed the target of the naturalization. We have in effect changed the subject, due to the same conflation described earlier: we were supposed to be identifying the property of *being right* and have instead offered up a *right-making* (but non-evaluative) property. At the very least, being right should be identified not simply with being happiness maximizing, but

with the *dispositional* property of *being such as to elicit relevant desires under relevantly idealized conditions* (cf. Smith 2004, 202); then the claim would be that acts have this complex property when they are happiness maximizing, which latter is right-*making* rather than rightness itself. That would be a better version of this functionalist approach.[14]

In any case, the more fundamental question we now need to address is whether even the best versions of naturalizing projects can suffice to give us everything we want in the way of normativity. I have tried to identify the more promising approaches, which perhaps fall under the category Parfit refers to as "soft-naturalism," since on these views normative claims give us distinct information not conveyed by mere attribution of properties such as being happiness maximizing (Parfit 2011, 364–369). As far as I can see, though, Parfit's critical arguments do not work against these views when they are formulated in the ways I've suggested: as we've seen, these more complex forms of naturalism can avoid worries about implausible trivialization of non-trivial claims, they can make sense of normative properties as distinct from right-making properties, and so on. Nonetheless, they remain problematic, and I now turn to developing these deeper worries.

7 Preliminary Doubts about the Resources of Naturalism to Account for Normativity

Here we come back to the issue of how much we want from an account of normativity. If we were prepared to drop categoricity we could offer a standard-based account that is just keyed to some stipulated end, such as social stability, or equality, greatest happiness, maximal freedom to pursue self-interest, or maybe the social cultivation and indulgence of rare creative genius. Given some such end, suitably specified, together with the natural facts and circumstances of life, there might well be objective practical standards. Ethical codes might be viewed more or less as social artifacts with end-relative functions, just like other artifacts. But none of this does anything to show that the end in question is the *proper* end, or the related standards the *correct* set of standards we all have reason to build our lives around.

Such an unambitious approach, however, runs the risk of securing naturalization at the cost of just changing the subject, missing ethical normativity in the rush to capture some naturalized use of normative concepts. If it were that easy to naturalize ethics, after all, Mackie would never have been

---

[14] Something similar should be said even about redness: it should not be identified simply with some microphysical property as such (a primary quality), but with the more complex relational, dispositional property (a secondary quality on the model of a Lockean power), which brings the relevant structure back in. In other words, the dispositional aspect should come in not merely as the *criterion* we use to identify the relevant natural property, but as the *nature* of that property itself.

driven to error theory. One might try to reduce the arbitrariness by appeal to conceptual constraints on what kinds of ends and standards can count as providing an ethical scheme—perhaps only those having to do with human goods and harms, or social-level functioning, say. But that doesn't help much because there are still so many sharply different possibilities, as noted above, and no sense so far that any of them has any real normative authority for us—any sense that someone would be making an objective mistake in shrugging off the stipulated end and associated standards. That makes it dubious that we're really capturing *ethics* with such minimal schemes, as Mackie emphasized. But even if he was wrong about that, and we cannot rule out such views just by appeal to the concept of ethics, many of us will still find them seriously deficient. Anyone who holds that ethics does essentially involve strong reasons to conform to a much more limited range of standards will obviously demand more of any naturalizing project.

Similar worries apply to parallel moves made within many functional role theories. Very little would be accomplished, for example, by a view focused on the role of causally explaining our use of the term or concept in question, such as 'right.' Such an approach might yield some natural property that plays that role, but so what? How could substantive normative questions about how we should live be settled for us by facts about how we use words?[15] Similar questions equally arise for more plausible appeals to functional roles. Consider a view (broadly along Cornell Realist lines) that takes the role relevant to rightness instead to be the promotion of social stability or flourishing, so that rightness is identified with the "property of acts that is causally responsible for their tendency towards social stability."[16] Such a view offers a naturalistic account of the features of action that play that causal role. But while this may be fine as far as it goes, it doesn't yet take us very far by itself, because it hasn't yet addressed the crucial matter of whether and why we all have strong *reason* to organize our lives around the role in question, such as the promotion of social stability or flourishing so conceived. If we seek an account of ethical properties such as *rightness*— not in some deflated form, but in all their normatively authoritative glory, as it were—then we haven't yet accomplished the task just by identifying a natural property that plays such a functional role. That might well be an important part of the story, but it does not provide a naturalistic account of the normative authority that at least many of us take to be integral to an ethical property such as rightness. This does not show, of course, that no further naturalist account could provide such a thing, and we will indeed look critically at two such attempts in the following sections. But it does indicate that the mere identification of natural properties playing

---

[15] Cf. Smith (1994, 32–34), who rejects such an approach as violating platitudes about disagreement in moral judgment and leading to an implausible, radical relativism.

[16] This characterization is taken from Smith (2004, 200), who goes on to reject such objective or "externalist" functional role views.

such intuitively ethically relevant functional roles (as promoting flourishing) does not yet get us there.[17]

This issue concerning the normative status of the role privileged by a given theory reveals a number of related puzzles. Granted, if we take as given a certain determinate conception of social stability or flourishing we might be able to identify a natural property that fills the role of best promoting it. But there are many possible conceptions of what a stable, flourishing society might look like—many alternative ways of weighing competing societal and individual needs, and of making tradeoffs among the elements of stability, freedom, welfare, equality, personal protections, and so on, even assuming some naturalized account of each.[18]

So what (naturalistically) determines the proper conception of the relevant goal of social stability or flourishing to begin with? To indicate some of the complications here, consider:

> Is the enslavement of a minority group by the majority a violation of the relevant goal even if most in the society are flourishing and there is so much control that it doesn't undermine social stability or prevent the society as a whole from meeting 'its needs'?[19] (And even if it did, would *that* be the proper explanation of what is wrong with such oppression?) Or alternatively, is the real goal egalitarian flourishing, so that everyone's flourishing is the target? Or is it instead the largely uninhibited pursuit of self-interest (constrained only as needed for social stability), even if this doesn't lead to flourishing for all?

If deep relativism is to be avoided, there must be *some fact of the matter* here as to the proper specification of the goal and corresponding role definitive of rightness. But now the question is: what purely *natural fact* serves to underwrite some one conception or specification of social stability and flourishing as the appropriate one to begin with? The worry here is not a first-order one over any particular view's extensional adequacy, or a methodological one over how we should work out answers to such questions: it is instead a worry about how any *purely natural fact* could do the relevant work. If there is a fact of the matter about which conception is the correct or normatively authoritative one, and this is to be a purely natural fact, then—to paraphrase Mackie—what *in the natural world* is *that*

---

[17] Smith (2004, 202), who himself embraces one sort of functional role view, agrees about the inadequacy of such objectively oriented or "externalist" versions, for very similar reason having to do with "justificatory potential".

[18] The last assumption is itself very contentious. It is far from clear that there is any purely naturalistic account of human flourishing, for example, that would show why that account rather than competing accounts is the correct one, picking out the sort of flourishing we all have most reason to pursue. On this, see in particular McDowell (1980). I discuss such issues further in FitzPatrick 2000, FitzPatrick 2008.

[19] For a naturalistic view focused on the meeting of societal needs, see Copp (1995).

fact supposed to consist in? Indeed, the question goes even deeper: what sorts of natural facts would determine that social stability and flourishing is even the proper goal (and promotion of it the relevant role) to begin with—as opposed to a Nietzschean one with very different priorities, for example? Or similarly, what natural fact is supposed to *constitute* the fact that some particular goal (under some particular conception) *merits* each person's special concern and allegiance (over various rivals), even if she doesn't happen to care about it, so that she has reason to conform herself to standards of rightness associated with the promotion of that goal?[20] Functional role accounts—at least versions focused on objective, non-psychological roles, such as those considered in this section—do not themselves provide an answer.[21]

We might, of course, imagine certain facts about human dignity or inviolability or equal worth settling such matters in a particular direction, perhaps in favor of liberal egalitarianism as against an egoist or Nietzschean ideal. But that thought makes sense precisely because such facts *remain evaluative or normative*, and have not been reduced to something else—just as nonnaturalists insist they should not be. The question is how any plainly *empirical* facts could do all the work of fixing and constituting the answers to these fundamental questions. Social scientists might tell us how various features of actions would make them stack up in light of various standards, by exploring their causal tendencies. It is hard to see, though, how they could discover some set of natural, objective facts that would constitute the more basic *normative* facts I've flagged about the standards or goals and roles themselves. Perhaps, though, naturalist philosophers can still capture such facts by appealing to some sort of complex function of partly *psychological* facts. I turn now to that.

8 Further Doubts: Psychologically-Oriented Naturalist Views

8.1 Internalist Realism and Naturalized Rationality

Naturalizers have available to them some initially plausible lines of argument here. One (discussed in the next sub-section) is to offer a naturalistic account of *reasons* in psychologistic terms, and then to relate the proposed ethical standards or roles to such reasons, thereby capturing the normative authority of those standards or role-related goals.[22] Another would be to provide an explanatory role account more along the lines of Smith's "internalist realism," specifying the role in terms of a relation to rational desire.

---

[20] Again, cf. Smith (2004, 202) for similar worries about "externalist explanatory role naturalism."

[21] A psychologically-oriented functional role view is considered in the next section, and a psychologically-oriented account of reasons, which could be used to supplement the functional role views considered in the present section is discussed after that.

[22] See Schroeder (2007) for the claim that practical normativity can be captured via a naturalized theory of reasons.

If the role played by the natural property is that of "eliciting desire under certain idealized conditions of reflection" (Smith 2004, 202), this might seem to get around the shortcomings I've described for other functional role views—assuming there *is* some such property. For surely, one might think, we *should* care about the property we *would* care about if we were informed and rational. There might not seem to be any room for an open question argument here, for example. There are, however, reasons to be skeptical that any such move will be adequate.

Smith's internalist naturalism sounds very plausible because of its appeal to *rationality*, which sounds like a promising way to capture normativity. But this promise is in tension with the commitment to a broadly naturalized conception of rationality—a feature that can be expected to be shared by other naturalists pursuing this sort of strategy. What we're talking about are "idealized conditions of reflection" wherein our desires are made to form a maximally coherent and unified set through ethically-neutrally specifiable procedures applied in the presence of empirical information. The question, then, is whether this conception of rational desire captures what it needs to: is the concept of *rightness* really just concerned, as an a priori matter, with the object of such desire (with respect to a feature of actions)? Or to put it another way, if there were such a concept that fit that role, would it really be the one we have most reason to give pride of place in our practical deliberations?

I deny that it is *a priori* that talk of rightness is talk of whatever feature of action (if any) plays that role, because we may reasonably be concerned with something deeper than that in seeking the right. We might think with Aristotle, for example, that there are real values determinative of what constitutes acting well that we are not guaranteed to grasp simply by having empirical information and cleaning up our given desires with ethically neutrally specifiable tools to make them more coherent and unified. Coming to understand what acting well involves, and to desire accordingly, may require having the right kinds of formative ethical experiences and training as well—something not specifiable in ethically neutral terms (McDowell 1995). If this is right, then we should reject the idea that a person can "figure out what is right by just thinking about the circumstances of action" (Smith 2004, 203), and instead insist that rightness is something beyond what can be guaranteed to be graspable by just any rational person (regardless of ethical background and experience) who thinks about the circumstances of action. The correct answers to questions about rightness will not be determined by facts about the objects of naturalistically idealized desires (which desires are not, after all, "immune to rational criticism").

Now these Aristotelian concerns might of course be mistaken. But they are in any case intelligible thoughts pertaining to ordinary talk of rightness, not straining its meaning or violating platitudes. So even if we are mistaken in thinking that there is more to rightness than what Smith allows, we are certainly not just "misusing the word 'right'," as he suggests

(Smith 2004, 203). The issue here is a substantive one. Obviously the concerns I've raised won't be shared by those who start with a commitment to naturalism and are prepared to settle for whatever values, if any, we can derive via naturalistic operations on desires. But those of us who instead start from robust commitments to certain values, and doubt that naturalized rationality is enough to get us to desire in accordance with them, will justifiably resist any view telling us that correct answers to questions about rightness are ultimately a function of facts about naturalistically idealized desires.

Smith might try to defuse these worries by granting that idealized desires don't fix the correct answers to questions of rightness if we're talking just about an *individual's* idealized desires as such. That narrow focus would indeed make morality implausibly arbitrary and dependent on contingent desires, allowing in principle for too much relativism and leaving room for open question arguments. On his view, however, the real focus is on facts about what all rational agents would *convergently* desire with respect to their actions. If there were facts about what all rational agents would desire their acts to be like, then the convergence must be stemming from rationality, rather than from arbitrary contingencies of desires. And this, he thinks, makes the object of this idealized desire a plausible candidate for rightness (Smith 2004, 204–206).

In fact, however, the same worry remains. I suppose if there were some natural property that all possible rational agents would desire their acts to have, regardless of the desires they started out with, it would be hard to deny that we should care about it. And perhaps there is such a property. The property acts can have of being effective in achieving one's ends, for example, plausibly fills this role: we would presumably all want our acts to have that feature. But if we're sympathetic to the Aristotelian perspective I've sketched, we won't find much pull toward thinking that this sort of device will get us all the way to rightness. We would instead expect convergence only on some 'lowest common denominator' that doesn't depend on having the right starting points in experience and desire, as with the minimal property of efficacy just mentioned. Even if it goes a bit further than that, why should we allow rightness to be held hostage to this sort of limitation, which might still be significant?[23] The only claim in this neighborhood we should obviously grant is that rightness is the feature we would desire our actions to have if our desires were not only coherent, unified, and empirically informed, but also *ethically informed* by appropriate experience of relevant values (FitzPatrick 2004). And in that

---

[23] We could, of course, build more into the notion of rationality, smuggling in evaluative assumptions or criteria (e.g., that it's part of rationality to respect every person's basic worth) that would make it more plausible to link rightness to rationality. But while I have nothing against relying on evaluative assumptions, we can't smuggle them into an account of rationality and then still claim that the account is fully naturalistic.

case we may as well just appeal directly to the values that are doing the real work and leave the desire-based naturalizing project behind.

Even more concerning than the potential limitations imposed by the convergence condition under naturalistic idealization is the implication that it could very easily turn out that morality is "a mere illusion"—that is, if there wouldn't be sufficient convergence of people's desires concerning some natural feature of actions (Smith 2004, 205). Obviously no realist, naturalist or nonnaturalist, can completely deny the theoretical possibility that he may be wrong about morality—that it could turn out to be an illusion after all, despite our apparent understanding to the contrary. But it is one thing to acknowledge this abstract possibility (which should go without saying), and quite another to structure one's theory such that moral truth is made contingent on something like convergence of desire under naturalistic idealization, and so to embrace the position that if this condition turns out not to be met then one will accept nihilism. No one who starts with the strong moral commitments mentioned earlier should rush to embrace an account that makes morality hang in that particular balance. This is especially true if we are sympathetic to the Aristotelian perspective that makes robust convergence under naturalized idealization appear rather unlikely.

I can't say with great confidence whether or not there would be a robust convergence, but I can say that if there turned out not to be then I wouldn't surrender to nihilism and conclude that it turns out to be an illusion, for example, that it is morally wrong to stone gays or rape victims, or to deprive girls of education, or to make voting rights contingent on race, and so on. We know moral facts like that with far greater certainty than we know anything about what human beings would all desire under ethically neutrally specifiable idealizing conditions regardless of their upbringing and ethical training. I do not, then, feel the "pull of nihilism" that Smith speaks of experiencing in pessimistic moments, which would make such a naturalistic gamble seem plausible. The idea of convergence may have a place, but the only obviously meaningful convergence would be one arrived at *through a common appreciation of irreducibly evaluative or normative properties and facts*—which properties and facts can plausibly be accommodated only within a nonnaturalist framework.[24]

---

[24] Isn't the nonnaturalist vulnerable in an exactly parallel way, having made the existence of moral truth contingent on the existence of *nonnatural properties and facts*, which are on at least as shaky ground as Smith's convergence? No, because the two cases are not parallel. They would be if the nonnaturalist abandoned the appeal to irreducibly evaluative properties and facts and instead offered a reductive account of moral truth in terms of some metaphysical state of affairs involving nonnatural properties or facts, analogous to the psychological state of affairs to which Smith appeals. In that case, a discovery that such a metaphysical state of affairs did not in fact obtain (the nonnatural stars fail to line up in the required way!) would be parallel to a discovery that there is no ideal convergence of desires, and there would be parallel trouble. But the nonnaturalist does not appeal to nonnaturalist metaphysics in this way. Moral truth is not cashed out in *non*-moral metaphysical terms any more than in psychological

8.2 Naturalized Reasons

The worries about trying to capture the normative authority of a proposed set of standards by connecting it up to a naturalized theory of reasons are similar, and I will conclude with them. The most popular approach to naturalizing reasons is a neo-Humean one, again attempting to extract normativity from psychology. Mark Schroeder (2007) articulates the motivation for such a view using a simple example involving two agents and a party: the fact that there will be dancing at the party is a reason for Ronnie to go, because he likes dancing, and it is a reason for Bradley not to go, because he hates it. This example is supposed to illustrate the fact that some reasons are "explained in some way by the psychological features of the agents for whom they are reasons," such as liking or hating dancing; and this is supposed to lead, in a Humean vein, to the "natural idea" that "all reasons are at some basic level pretty much like Ronnie's and Bradley's" in this respect (2007, 1–2). Similarly, the idea is that it is "uncontroversial that reasons are sometimes explained by psychological features," which leads to the thought that:

> there can be a unified explanation of why people have the reasons they do only if all reasons are explained by this kind of psychological feature. . . . If there is any unified story about [what reasons are and where they come from], and it sometimes works by appealing to psychological features of some kind, the Humean thought is that it must always work by appealing to psychological features of that kind. (2007, 1–2)

While I agree that the fact that there will be dancing at the party is both a reason for Ronnie to go and a reason for Bradley not to go, I reject all of the Humean claims Schroeder makes about the case, and deny that it provides any real motivation for a Humean view.

To begin with, the case does not show or even suggest that some reasons are explained simply by the psychological features of agents. It shows only that the *difference* in agents' psychologies can sometimes make a *difference*

---

terms: it is just the holding of irreducibly moral facts. The nonnaturalist metaphysics comes in differently: it is simply a theoretical posit when it comes to *characterizing the nature of these irreducibly moral facts*. We posit nonnatural properties and facts not to play the sort of role psychological convergence plays for Smith, but simply as our best way of understanding the metaphysical nature of the irreducibly evaluative or normative. Of course, we could well be wrong about this: indeed, I am the first to grant that we should be far more confident in our central first-order ethical views than in any piece of speculative metaphysics such as normative nonnaturalism. But if I were led on metaphysical grounds to reject the existence of nonnatural properties and facts this would not lead me to embrace *nihilism*: I would simply abandon metaethical nonnaturalism at that point and look for a better theory—a more satisfactory metaphysical characterization of normative properties and facts. This is very different from Smith's acceptance that if ideal psychological convergence fails then nihilism wins. (Thanks to an anonymous referee, and to Jon Tresan, for pressing this objection.)

to the reasons they have (Scanlon 2014, 47–50). Even in this simplest case, then, it is far from clear that the psychological facts are directly grounding the reasons. The other possibility is that the psychological facts are merely contributing elements with respect to some larger factor that generally grounds reasons; and that factor may not turn on such psychological elements at all in many other cases, *contra* the Humean generalizing thought. For example, on a *value-based view of reasons*, one might hold that for *R* to be a reason for *X* to do *A* is for *R* to bear positively on the unqualified goodness of *X*'s doing *A*, so that the truth of *R* helps explain why *X*'s doing *A* will constitute acting well. This is what an Aristotelian might say, taking the proper aim of practical reason to be not desire-satisfaction but *eupraxia*. On this view, the psychological differences between Ronnie and Bradley are difference-makers for their reasons, but only because they happen to be among the relevant circumstances that determine what would constitute acting well or good action for each of them—not because their desires are directly grounding different reasons. The point is just that it is part of acting well to pursue innocent pleasures, but not to torture yourself pointlessly, which is why the psychological differences make a difference to the reasons the two agents have. There is no suggestion that such psychological factors play a similar role in *all* cases, let alone that they fully ground reasons in *any* case.

The value-based view thus explains cases like Ronnie and Bradley no less than the desire-based view does, and it has just as much explanatory unity in relating all reasons to the relevant goodness of action, again *contra* the passage quoted above. Moreover, there is no implication that goodness of action is always a function of our desires or likes or dislikes. On the contrary, many of us will think that all sorts of other considerations, such as fairness, justice, and other people's needs can bear directly on what will constitute acting well for us, and therefore on what reasons exist for us, quite independently of such facts about our psychologies. If so, then on this view we should not suppose that just because desires play a (limited, contingent) role in the party example they likewise explain all reasons. They don't even fully explain those simple reasons, let alone all the rest, but are just contributing factors in some cases, where they are among the circumstances bearing on goodness of action. So the whole motivation for the neo-Humean project depends on a construal of cases like the party case that those sympathetic to a view that bottoms out in evaluative facts about action will reject from the start (as will many who take reasons to be irreducible and fundamental, as noted below).

Now this is not yet to address head-on the *naturalism* in Schroeder's view, since the above objection to the claims about the motivation for Humeanism could be shared even by naturalist value-based theorists of reasons; nor is this a form of objection that all nonnaturalists will share, since many nonnaturalists take reasons to be more fundamental than value, thus rejecting my suggested value-based view—though they will equally

reject the Humean grounding of reasons in desires and will provide their own analyses of cases such as Ronnie and Bradley (Scanlon 2014). But it is still relevant to the critique of naturalism because the form of naturalism we're considering in this section is one that cashes out reasons in terms of psychology, and it is therefore worth starting by undermining the appearance that commonsense thoughts about cases such as Ronnie and Bradley push us firmly in a psychologistic—and hence *naturalistic*—direction. The possibility of the sort of value-based view sketched above shows that they do not. And precisely because there are serious doubts about whether non-psychologistic naturalistic accounts of value can account for its normative authority or reason-giving force, as discussed earlier, those of us inclined to a value-based view will also be pushed toward a nonnaturalist version bottoming out in irreducibly evaluative facts and properties. So it makes sense to appeal to plausible intuitions about value-based reasons in seeking to undermine the pull of psychologistic naturalism about reasons.[25]

To stick with the value-based view for a moment: if we find such a view plausible—at least as contrasted with a desire-based view—we shouldn't even accept the common idea of *instrumental normativity* or *reasons*, if this is construed as some special kind or normativity or class of reasons grounded in desires or ends. My resistance here is again based on normative considerations. It is mysterious, to say the least, why we should think that someone could generate a real, positive normative reason to engage in gay bashing, say, just by having nasty enough desires or willing a nasty end. Any view that implies that genuine normativity can arise from mere psychology in those ways seems just to be turning the idea of normativity into something else, such as motivation or instrumentality.[26] Why should

---

[25] None of this is intended to suggest an argument against nonnaturalist reasons-fundamentalism such as Scanlon's (2014), though I believe such a view is problematic for other reasons. The point here is simply to undermine the motivation for the Humean psychologistic turn in accounting for reasons. Indeed, the spirit of the critique that follows is very much shared by nonnaturalist reasons-fundamentalists. They will equally resist the idea, for example, that mere desires as such give rise to reasons, or that reasons depend in a general way on the presence of relevant desires, even if they don't defend this resistance with a more basic appeal to value.

[26] Reducing the property of being a reason to the mere property of explaining why an act is *instrumental* to the satisfaction of a desire will saddle one with "too many reasons," among other things—an objection Schroeder recognizes and addresses by claiming that the bizarre reasons generated, such as the reason to eat one's car, are vanishingly weak. (Note that this is distinct from the problem with normatively implausible reasons, such as reasons to engage in gay bashing, based on *objectionable desires*. It is no help there being told that such reasons are just very weak; nor is it clear how they can be guaranteed to be weak in people with corrupt enough desires that will affect also their reasons for assigning weights to other reasons.) The value-based view, by contrast, isn't committed to such reasons at all. While we acknowledge the instrumentality claim, we can reject the idea that this fact *bears positively on the goodness* of eating one's car. It does not, any more than the fact that burning down the house will roast the pig bears positively on the goodness of burning down the house. Claims about goodness of action are more holistic than mere claims about instrumentality, which are atomistic and

mere psychology, divorced from relevant facts about value, be thought to generate positive normative force—even weak force—in favor of doing something, however stupid or evil?[27] There seems to be something akin to a category confusion here. Likewise, any view that implies that someone may well lack any strong reason to refrain from gay bashing if only his psychology turns out to be immoral enough seems to be unacceptably watering down ethics. It holds the normative significance of ethics hostage to the psychological contingencies of agents in a way that robs ethics of its proper normative force. One might speculate about the possibility of universal desire-based reasons associated with morality, but as before, I reject that naturalistic gamble: I'm far more confident in the normative force of core moral requirements than I am that a desire-based theory would deliver that result.

A theory such as Schroeder's does better than some by recognizing that the *concept* of a reason is *irreducibly normative*. It therefore does not falsely imply, for example, that someone would be conceptually confused if she continued to question whether something is a reason (or denied that it is) even after having been told the facts Schroeder highlights about desires, their objects, and instrumental relations between potential actions and those objects. The real problem, however, is that it does still imply that facts about reasons are entirely fixed by (or are a function of) such facts: there is nothing else bearing on the correct answers to questions about reasons. (Compare: act utilitarianism likewise implies that facts about rightness are entirely fixed by facts about an act's consequences for overall utility: there is nothing else bearing on the correct answer to questions about rightness.) On such a view, if we continued to hold out, questioning whether something really is a reason (or denying that it is) even after having been told these psychological and instrumental facts, then we would be making a fundamental *mistake* in doing so, because the facts about reasons are determined entirely by such facts and there is no room for anything else to make a difference to what reasons there are. It is precisely that implication that is deeply problematic for anyone who shares the normative intuitions I have emphasized.

Suppose my doing something to someone will promote the object of some desire of mine, and this is partly explained by the truth of some proposition

isolated from broader considerations. Mere instrumentality to roasting the pig does nothing to support the *goodness* of burning down the house.

[27] There can, of course, be cases of *derivative* reasons even in cases involving evil desires. If someone's not engaging in gay bashing will result in his having a terrible migraine, for example, then avoiding the migraine seems to provide some reason for acting that is simply overridden by the reasons against it. Here the derivative end is the good one of avoiding the headache, as opposed to the evil one of causing gays to suffer, and any normativity is coming from the former, not the latter. This is similarly the only sort of non-disjunctive, positive normativity generated by the instrumental principle's call to *avoid inconsistency* in the case where the agent for some reason cannot drop a bad end, and it is again easily overridden by other reasons, such as reasons to avoid injustice.

R. On Schroeder's view, this determines the correct answer to the question whether *R* is a reason for me to do this thing, because *R's* explaining why my doing something would fulfill a desire of mine *just is* what it is for *R* to be a reason for me to do it. I reject that identification, however, and this is (in part) because I do not accept that such facts determine the answers to such normative questions: other things matter to the existence of reasons, such as the goodness of the desires (or more relevantly, the goodness of relevant ends and of the means in question). Nor can I accept that the correct answers to questions about what we have most reason to do are likewise fixed by facts about higher-order reasons for weighing other reasons, where these facts are again understood just to bottom out in facts about desires.

One might understandably go for such a view if one believed that there are no evaluative facts about ends and actions of the sort that figure into the value-based view—though one might equally conclude that in that case there aren't really any normative facts about reasons either, but just various psychological facts. But those of us who do posit robust evaluative facts about ends and actions have good cause to resist any such psychologistic treatment of reasons. Again, notice that the worries raised in this section are not just against the idiosyncrasies of one particular view, such as Schroeder's: they are very broad ones that will plausibly apply in some form to any psychologically-oriented naturalistic account of reasons.[28] Of course, those who hold a naturalist value-based theory of reasons will agree with this part of the critique while still holding on to naturalism. But as discussed earlier, there are grounds for being skeptical about the prospects for naturalized accounts of value when it comes to fully capturing its normative significance for us: that was precisely Smith's criticism of "externalist" functional role views, such as those focused on social stability, and what led him to turn instead to idealized desires. Those earlier worries, then, equally raise doubts about whether a naturalized view of value could do the work of grounding a satisfactory theory of reasons: it seems instead to be in need of supplementing with a distinct theory of reasons. The line of resistance I have offered against views such as Schroeder's therefore seems to support either a nonnaturalist reasons fundamentalism or a nonnaturalist value-based view that posits irreducibly evaluative properties and facts—such as facts about human worth and dignity, and facts about their meriting our concern in practical deliberation.

9 Conclusion

I have tried to show that while many are led to embrace naturalizing accounts of normativity, including ethics, out of a prior commitment to

---

[28] Perhaps neo-Kantian constitutivist accounts avoid the problems raised for Humean views, but as noted earlier, they suffer from equally serious problems, as I have argued elsewhere.

metaphysical naturalism, and find resistance to naturalizing ethics puzzling (at least once nihilism is rejected), there are in fact very good reasons why one might be led to embrace nonnaturalism. This has nothing to do with an odd penchant for "spooky" additional facts and properties that are what really matter or do the real justificatory work in ethics. It is instead rooted in a conviction that we need to posit irreducibly normative—and hence nonnatural—facts and properties to make adequate sense of certain normative metafacts or metaproperties concerning the normative significance of natural facts and properties.

My claim has been that naturalizing accounts seem to fail in this regard, at least if we refuse to bite certain bullets or to deflate normativity in various ways. Some such accounts just fail altogether to capture the normativity they're after by identifying normative facts or properties with plainly non-normative ones that instead play a different role, such as being good-making or right-making facts or properties. Others are more promising, but they still turn out to be limited in what they can deliver, especially if we are careful not to smuggle in any nonnaturalized normative assumptions. Indeed, some will have distinctly counterintuitive implications. I have not, of course, shown that such views are therefore false. Whether that is the case or not depends partly on whether the intuitions I have relied upon—Aristotelian thoughts about ethical values that outrun naturalistically idealized desires, or intuitions about non-relative ethical standards or objective-value-dependent and categorical reasons—are correct. What I do claim to have shown is that anyone who takes those intuitions seriously at least has substantial reason to be skeptical about the prospects for naturalizing projects in the broad neighborhood of those considered here, and should be on the lookout for similar shortcomings in other proposals. I hope at the same time to have illuminated the respectable motivations for normative nonnaturalism, as a corrective to the caricatures it often receives.

William J. FitzPatrick
E-mail : william.fitzpatrick@rochester.edu

References:

Copp, David. 1995. *Morality, Normativity and Society*. Oxford: Oxford University Press.

Dancy, Jonathan. 1993. *Moral Reasons*. Oxford: Blackwell.

Dancy, Jonathan. 2004. "On the Importance of Making Things Right." *Ratio* 17 (2): 229–37. http://dx.doi.org/10.1111/j.1467-9329.2004.00250.x.

Dancy, Jonathan. 2006. "Nonnaturalism." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 122–45. Oxford: Oxford University Press.

Dreier, James. 2004. "Meta-ethics and the Problem of Creeping Minimalism." *Philosophical Perspectives* 18 (1): 23–44. http://dx.doi.org/10.1111/j.1520-8583.2004.00019.x.

FitzPatrick, William. 2000. *Teleology and the Norms of Nature*. New York: Garland.

FitzPatrick, William. 2004. "Reasons, Value, and Particular Agents: Normative Relevance Without Motivational Internalism." *Mind* 113 (450): 285–318. http://dx.doi.org/10.1093/mind/113.450.285.

FitzPatrick, William. 2005. "The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity." *Ethics* 115 (4): 651–91. http://dx.doi.org/10.1086/430750.

FitzPatrick, William. 2008. "Robust Ethical Realism, Non-Naturalism and Normativity." In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, Vol. 3, 159–205. Oxford: Oxford University Press.

FitzPatrick, William. 2011. "Ethical Non-Naturalism and Normative Properties." In *New Waves in Metaethics*, edited by Michael Brady, 7–35. London: Palgrave MacMillan.

FitzPatrick, William. 2013. "How Not to Be an Ethical Constructivist: A Critique of Korsgaard's Neo-Kantian Constitutivism." In *Constructivism in Ethics*, edited by Carla Bagnoli, 41–62. Cambridge: Cambridge University Press.

Foot, Philippa. 2001. *Natural Goodness*. Oxford: Oxford University Press.

Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Gibbard, Allan. 2006. "Normative Properties." In *Metaethics After Moore*, edited by Terence Horgan and Mark Timmons, 319–38. Oxford: Oxford University Press.

Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford: Oxford University Press.

Korsgaard, Christine. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Korsgaard, Christine. 1997. "The Normativity of Instrumental Reason." In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Gaut. Oxford: Clarendon Press.

Korsgaard, Christine. 2009. *The Constitution of Agency*. Oxford: Oxford University Press.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.

McDowell, John. 1980. "The Role of *Eudaimonia* in Aristotle's *Ethics*." In *Essays On Aristotle's Ethics*, edited by Amelie O. Rorty, 359–76. Berkeley, CA: University of California Press.

McDowell, John. 1995. "Might there be External Reasons?" In *World, Mind and Ethics*, edited by J.E.J. Altham and Ross Harrison, 68–85. Cambridge: Cambridge University Press.

Paakkunainen, Hille. 2014. "How Constitutivism Works." Manuscript, presented at the Central Division Meetings of the American Philosophical Association, symposium session on Naturalizing Normativity.

Parfit, Derek. 2011. *On What Matters*. Vol. 2. Oxford: Oxford University Press.

Plantinga, Alvin. 2010. "Naturalism, Theism, Obligation and Supervenience." *Faith and Philosophy* 27 (3): 247–72. http://dx.doi.org/10.5840/faithphil201027328.

Ridge, Michael. 2014. *Impassioned Belief*. Oxford: Oxford University Press.

Scanlon, Thomas. 2014. *Being Realistic About Reasons*. Oxford: Oxford University Press.

Schroeder, Mark. 2007. *Slaves of the Passions*. Oxford: Oxford University Press.

Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.

Smith, Michael. 2004. *Ethics and the A Priori*. Cambridge: Cambridge University Press.