UNIVERSITY of York

This is a repository copy of *Fittingness : The Sole Normative Primitive*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/81428/

Version: Submitted Version

Article:
Chappell, Richard Yetter orcid.org/0000-0003-3322-4935 (2012) Fittingness : The Sole Normative Primitive. Philosophical quarterly. pp. 684-704. ISSN 0031-8094

https://doi.org/10.1111/j.1467-9213.2012.00075.x

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Fittingness: The Sole Normative Primitive

(Forthcoming in *The Philosophical Quarterly*)*

Richard Yetter Chappell†
Princeton University

April 6, 2012

Abstract

This paper draws on the 'Fitting Attitudes' analysis of value to argue that we should take the concept of *fittingness* (rather than *value*) as our normative primitive. I will argue that the fittingness framework enhances the clarity and expressive power of our normative theorizing. Along the way, we will see how the fittingness framework illuminates our understanding of various moral theories, and why it casts doubt on the Global Consequentialist idea that acts and (say) eye colours are normatively on a par. We will see why even consequentialists, in taking rightness to be in some sense *determined by* goodness, should not think that rightness is conceptually *reducible* to goodness. Finally, I will use the fittingness view to explicate the distinction between consequentialist and deontological theories, with particular attention to the contentious case of Rule Consequentialism.

*Please cite final published version once available.

†For helpful feedback on this paper, I'd like to thank Ryan Cook, Rob Lawlor, Errol Lord, Philip Pettit, Doug Portmore, Andrew Reisner, Peter Singer, Michael Smith, Nic Southwood, Helen Yetter Chappell, and audiences at Princeton, ANU, and the *PEA Soup* blog.

1 Introduction

Consequentialists have traditionally followed Moore (1903a) in taking the concept of *value* as their normative primitive. This metaethical view may be accompanied by the assumption that axiology exhausts normative ethics: i.e., it's a substantive question what's valuable (good or bad), but there's no *further* question of 'right' and 'wrong'. Some recent theorists have embraced the implication that we should not privilege acts as uniquely subject to normative evaluation in terms of 'rightness'. Instead, they claim, any normative status that's applicable to acts, can just as well apply to other possible objects of evaluation, from eye colours to climates. Call this the **Parity Thesis**. Such parity might be achieved by either shrinking or extending the domain of 'rightness'. These two possibilities are reflected in Scalar and Global Consequentialisms, respectively. Global Consequentialism defines the 'right' x (for any category of evaluand x, be it acts, climates, or whatever) as the best x of those available (Pettit and Smith 2000). It thus indiscriminately extends 'rightness' talk to all evaluands. Scalar Consequentialism, on the other hand, does away with 'rightness' altogether, and merely has us evaluate acts (like anything else) as more or less good.[1]

[1] At least, this is one natural form that Scalar Consequentialism might take. Call it *Evaluative Scalar Consequentialism*: the view that *all* we can say about actions is that they have more or less value, and there's no other normative claims — not even about, say, reasons for action — to make at all. (I see this as a terminological variant of Global Consequentialism, which will be my main foil for the remainder of the paper.) On the other hand, we can imagine a version of the view which, despite rejecting the *binary* distinction between 'right' and 'wrong', nonetheless allows that we can indeed make normative claims about choiceworthiness (or reasons for action) over and above evaluative claims, it's just

An alternative tradition follows Brentano (1902) and Sidgwick (1907)[2] in understanding value in terms of a more fundamental notion of *desirability* or what we have (fitting) reason to desire.[3] In this paper, I make the case for taking this notion of *fittingness* as the fundamental concept for use in normative theorizing. The Moorean's value claims can still be captured within the fittingness framework as a matter of the object's being *fit to desire*. But by appealing to the notion of 'fittingness' more generally, we expand our theory's expressive power. Not only can we talk about value (what's fit to desire), but also what's fit to believe, and — more importantly for our purposes — what it is fitting (right)[4] to *choose to do*. This conceptual framework allows us to make substantive new claims about the rightness of acts, that

that this further normative status is also a matter of degree. This more sophisticated view — call it *Normative Scalar Consequentialism* — is compatible with my rejection of the parity thesis.

I'm not certain which version Norcross is advocating in his (2006), for while he does speak of 'reasons for action', it isn't clear whether or not he thinks such claims are conceptually reducible to value claims.

[2] Andrew Reisner pointed out to me that it's controversial to what extent we should characterize Sidgwick as a fitting attitudes theorist — Moore himself did not think of Sidgwick as such (Moore 1903b). Nonetheless, I follow Smith (2010) in finding Sidgwick's talk of 'desirability' and 'rational ends', e.g. in section I.ix.3 of his (1907), at least highly suggestive.

[3] Scanlon (1998) and other deontologists understand value more broadly as warranting various kinds of pro-attitude (e.g. admiration, respect). For simplicity, I will focus exclusively on the consequentialist conception of value as desirability.

[4] Admittedly, the notion of fitting choice does not exactly correspond to the ordinary deontic notions of obligation or permissibility. In particular, as noted in fn.1, I don't mean to rule out the kind of 'normative scalar consequentialism' that treats moral rightness as coming in degrees, with no sharp line between those acts that are sufficiently fitting or justified to qualify as 'permissible' and those that aren't. For current purposes, the relevant similarity to our ordinary talk of 'right action' is just that we're talking about a kind of normative status that applies distinctively to one's choice of action. For more on the fuzziness of 'rightness'-talk, see Lawlor (2009, Appendix A).

are not merely disguised claims about value restated in different words. We will also see that the fittingness view provides a more ecumenical framework for normative theorizing, allowing logical space for both consequentialist and deontological theories to emerge.

1.1 Conceptual Frameworks

To get a firmer grip on the notion, we may understand a 'conceptual framework' to be a set of interrelated concepts, with some specified as primitives, and others defined in terms of the primitives. The conceptual framework for a domain is, in effect, the expressive toolkit through which we can make first order claims on some topic. In the case of the normative domain, I suggest three desiderata against which we may assess candidate conceptual frameworks:

1. It must provide adequate conceptual resources for us to *express* any expressible normative truth.[5]

2. It should, so far as possible, remain neutral on the disputes of first-order normative ethics.

3. It should reflect and illuminate any natural '*joints*' in the ontological structure of the normative domain (insofar as the normative domain is

[5] Of course, most normative claims involve the use of non-normative concepts too. So a more precise statement of the first desideratum is that the conceptual framework for our normative domain must provide adequate conceptual resources for expressing the *normative parts* of any expressible normative truth. It may require supplementation by non-normative conceptual frameworks in order to express the non-normative parts of a normative claim. Hopefully the intended idea is clear enough.

structured).

These desiderata are motivated by the philosophical role that we are seeking to fill. We are looking for a conceptual framework within which to conduct our first-order normative inquiry. A candidate conceptual framework will be of little use to us if we cannot use it to express the answers we seek. Gross non-neutrality in a framework similarly precludes using it for this purpose, insofar as it simply presupposes answers to the questions we wished to ask. The third desideratum is less strictly essential, but clearly a nice feature to have if possible. For example, it may help illuminate which first-order debates are genuinely significant, whereas a framework with (e.g.) redundant concepts risks encouraging terminological disputes and other confusions.

1.2 Roadmap

The structure of this paper is as follows. In section 2, I motivate the fittingness view in relation to other positions in the literature, with particular reference to its expressive power. Section 3 pits the fittingness framework against value primitivism with respect to the third desideratum. To anticipate: the fittingness view implies that there is a distinct normative status corresponding to each kind of judgment-sensitive 'rational output' that can be assessed as fitting or unfitting to the situation, e.g., beliefs, desires, and actions. So, in particular, there's a distinctive normative status that is applicable to actions, with no analogue for mere evaluands like eye colours. I offer several reasons to think that this accurately reflects an important 'structural'

difference between judgment-sensitive and non-judgment-sensitive evaluands, thus arguing against the value primitivist's Parity Thesis, and by extension, against Global Consequentialism.

Section 4 uses the above asymmetry in the fittingness framework to defend Act Consequentialism against the Global Consequentialist's charge that, in rejecting the Parity Thesis, we are "privileging" acts in a theoretically objectionable way. Whereas advocates of Global Consequentialism have claimed that theirs is the view that Act Consequentialists were striving for all along, I argue that the reverse is true: Global Consequentialists who accept the fittingness view should be led to Act Consequentialism.

Section 5 defuses an important objection by showing how to understand the structure of deontological and rule consequentialist theories in terms of the fittingness view. Finally, in section 6, I consider the objection that facts about *fitting choice* may reduce to facts about *fitting desire*. I argue that, while the two may typically coincide, no such reduction is possible, at least not without collapsing the consequentialism-deontology distinction.

2 Reasons-Talk and Fitting Attitudes

To begin, let me clarify how Scanlonian 'reasons fundamentalism' (the view that takes *reasons* as the fundamental normative concept) fits into the dispute between the fittingness view and value primitivism. I see the fittingness framework as very much in the spirit of reasons fundamentalism. My worry about starting with reasons-talk is that in practice it seems to invite unnec-

essary ambiguity and confusion. Normative reasons are typically introduced as 'facts that count in favour' of an act or attitude, but there are two very different kinds of 'favouring' that might be invoked here.[6] This is seen, for example, in the distinction between 'epistemic' and 'practical' reasons for belief. Suppose that a prankster demon will reward me for believing that grass is purple. This fact might be said to 'count in favour' of the belief, in one obvious sense: it establishes that there is (instrumental) value to my having this belief — it would be a *fortunate* state for me to be in. But there's a very different sense in which evidence, e.g. the fact *that grass looks green*, 'counts against' believing grass to be purple. Evidence counts for or against belief in a proposition by speaking to whether the proposition is true and hence belief-worthy or *fit* to believe.

We can similarly distinguish value-based reasons and fitting reasons for *desire*, say if we would be rewarded for desiring that others suffer. Suffering is undesirable: it has qualities that 'count against' desiring it, in the sense of rendering such desires *unfitting* to their objects. But external incentives may 'count in favour' of such perverse desires, by rendering their possession instrumentally valuable.

In light of this ambiguity, it is unhelpful to take the notion of 'counting in favour' as primitive (in our normative theorizing). We must further specify

[6] This distinction is typically analysed in terms of 'object-given' versus 'state-given' reasons for attitudes. I avoid this terminology because there may be tricky cases where the object of the attitude is in some sense responsible for our having what should intuitively be characterized as a state-given reason. Cf. Rabinowicz and Rønnow-Rasmussen (2004).

that our conception of 'reasons' is that of counting in favour *in the sense of* rendering an attitude fitting, rather than merely fortunate. Otherwise, the inclusion of value-based reasons leads to the 'Wrong Kind of Reasons' problem (Rabinowicz and Rønnow-Rasmussen 2004) for analysing value in terms of reasons for desire. That is, an external incentive to desire suffering would then suffice to give you "reason to desire" suffering, though it clearly wouldn't make suffering itself good, thus undermining the identification of what's good with what we have reason to desire.

We avoid this problem if we start from the unambiguous notion of fittingness: external incentives to desire suffering don't make suffering itself any more desirable (fit to desire), after all. So rather than taking reasons directly as primitive, we do better to reconstruct the notion of 'reasons for desire' explicitly in terms of *desirability characteristics* — features in virtue of which an object is fit to desire.[7] This allows us to better avoid objections when it comes time to analyse value in terms of our normative primitive.

The unhelpful ambiguity likewise infects other general normative terms like 'ought'. There seems a sense in which you ought to believe/desire what's fitting, and another sense in which we can say that you 'ought' to believe/desire in whatever way would be best. It doesn't seem like there's any way to balance and combine these into a single normative conclusion;

[7] Rather than starting from the *features* that render an object fitting to desire, we might follow Smith (1994, 2010) in taking as fundamental the *ideally rational psychology*, and then understand reasons for desire (or desirability characteristics) as *features that would figure in the ideal agent's desires*. I mean to remain neutral on this.

rather, the two kinds of evaluations are giving different answers to different questions.[8] So rather than starting with general normative terms like 'reasons', 'counts in favour', 'ought', or the like, we should start from our understanding of the two more specific kinds of normative evaluation: the *fitting* (or rationally warranted)[9] and the *fortunate* (or instrumentally valuable).

We need to be able to make both kinds of evaluation, in order to make sense of the cases discussed above. But we can achieve this using fittingness evaluations alone. For whenever a mental state would be fortunate, that means that it would be desirable — fitting to desire — to possess that mental state. To illustrate: the *belief* that grass is purple is unfitting,[10] but it's

[8] As we'll see below, we can understand the two questions as (1) is p believable/desirable? and (2) is the belief/desire that p desirable? Even if the answers diverge (suppose the answers are 'no' and 'yes', respectively), an agent could be fit in both respects: that is, he could reasonably fail to believe p, whilst reasonably desiring to possess the irrational-but-fortunate belief. Cf. Parfit (2011, Appendix A).

[9] I don't here mean to take any stand on the issue of 'objective' vs. 'subjective' oughts, or how incomplete and misleading evidence affects the normative status of attitudes. I'm inclined towards a pluralistic stance: there's an 'objective' sense of fittingness that corresponds to objective reasons, or what we might call 'objective rationality' — roughly, what would be rational given full information, modulo the conditional fallacy (Shope 1978). I take there to also be various 'subjective' or 'evidence-relative' senses of fittingness, which correspond to ordinary evaluations of rationality or reasonableness for fallible non-omniscient agents. For simplicity, we may focus on fully-informed ideal agents, for whom 'objective' and 'subjective' norms coincide.

[10] Some people report finding it just as intuitive, to their ear, to apply the word 'fitting' to instrumentally valuable beliefs like this. I mean to use 'fittingness' as a technical term to pick out precisely that intuitive normative concept that assesses (e.g.) desires according to whether their objects are genuinely desirable, and beliefs according to whether their object merits belief, etc. In other words, it is the normative concept which encompasses 'epistemic' rather than 'practical' reasons for belief, generalized to apply to other attitudes as well. You should associate it with words like 'warranted', 'appropriate', 'right' or 'correct', and (on the negative side) 'crazy' and 'perverse'; **not** with 'useful', 'fortunate', 'important', or (on the negative side) 'disastrous'. I hope readers find it intuitively compelling that there is a general notion in this vicinity which is the common thread between credible beliefs, desiring what's desirable, and fearing the fearsome. There is likewise

perfectly reasonable to *desire* to have the belief, in light of the proffered reward. Likewise, though others' suffering is itself undesirable, it may be fitting to desire (for instrumental reasons) *that you (perversely) desire that others suffer*. In this way, value claims are analysable in terms of fittingness claims, allowing us to take the latter as our sole normative primitive without any cost to the expressive power of our normative theorising.

One might wonder whether we could just as well reverse this approach, and seek an analysis of fittingness in terms of value. But the prospects for such an analysis seem dim. For while there is an intimate conceptual connection between value and desirability, there is not any such obvious connection between value and (say) believability. The evidence may render belief in P fitting even when both P itself and *your believing of* P are unfortunate or disvaluable. We can even imagine worlds governed by truth-hating demons, where it is not even good *as a general rule* to believe what's supported by the evidence. So there seems little hope for analysing fitting attitudes in terms of value, the way that we can easily analyse value in terms of fitting attitudes.

This asymmetry suggests that the Fittingness framework has greater expressive power in the following sense: Any value claim can also be expressed in terms of fittingness, whereas it is not the case that all fittingness claims can also be expressed in terms of value. If we want to be able to make

something commonly *inappropriate* about crazy/incredible beliefs, perverse desires, and irrational phobias. We may capture this commonality by saying that each attitude is, in its own way, an *unfitting* response to its object.

both kinds of claims, using only one kind of primitive, then this suggests the primitive we should choose is that of *fittingness*, rather than *value*.

My remarks so far have suggested a strong presumptive case for taking *fittingness* as our normative primitive. This is a variation of a view held by many philosophers, including Scanlon (1998, 2009) and Darwall (2006). The original contribution of this paper comes next, as I develop the framework and demonstrate its theoretical payoff.

3 Rational Outputs and the Scope of Normative Theorizing

The scope of our normative theorizing is constrained by our normative primitives. Everything there is to say can be said using the primitive concepts — otherwise we'd need additional primitives. So, if Mooreans are right to take the concept of *value* as their sole normative primitive, it must be that the value facts in some sense exhaust the normative facts: there's nothing more to say — or, at least, nothing distinctively *normative* — once we've settled what's good and bad. In particular, this means that there is no normative status of 'rightness' or fittingness that applies distinctively to acts or choices, as opposed to mere evaluands like eye colours and the global climate. All we can say is that some acts are better or worse than others (in virtue of leading to better or worse outcomes), in just the same way that some eye colours may be better or worse than others (in virtue of leading to better or worse outcomes). If we want to apply the term 'right' to the instrumentally best acts, we may as well call the instrumentally best eye colour the 'right' one to

have too — though of course this is not to make a substantive new claim, but just to re-state the old evaluative claim using new words. In this way, value primitivists are naturally led to a kind of Global Consequentialism (Pettit and Smith 2000; Kagan 2000; Ord 2009), which we may examine here as a kind of case study of the Parity Thesis.

The essence of Global Consequentialism, as it interests me here,[11] is its symmetrical treatment of acts and other, non-judgment-sensitive objects of evaluation (what I call 'mere evaluands'). Pettit and Smith (2000, 122) write: "The crucial feature of global consequentialism is that it does not privilege any category of evaluand." Ord (ms) adds: "By encompassing all evaluands, rather than just the most prominent ones, it maximizes its expressive power while remaining nonarbitrary. Indeed, some systems of global consequentialism promise to be simpler than act-consequentialism itself, for by allowing everything to be morally assessed [in the same way], they no longer need an associated theory of acts." (p.4) What's distinctive about this view is its whole-hearted embrace of the **Parity Thesis** that the normative domain is symmetrically structured, such that actions and eye colours are on a par when it comes to the ways in which we can normatively assess them. There's nothing special about judgment-sensitive evaluands, on this view. To think otherwise, Global Consequentialists claim, would be to arbitrarily privilege

[11] One might offer a strictly weaker view that merely defends a global definition of *rightness*, whilst countenancing other normative evaluations — such as fittingness — that have non-global scope. But this would amount to a mere terminological variant of the act consequentialism that I defend in §4.

acts.

The value primitivist might try to escape this commitment to the parity thesis by constructing a new concept of 'rightness' out of a combination of the normative concept *value* plus the non-normative concept *can* (as it figures in the slogan, 'ought implies can').[12] So, for example, to call something 'right' might just be to say that: (i) the agent in question *could choose to do it*, and (ii) this choice would produce at least as much value as any other choice the agent could have made. Assuming that actions are just those things that an agent 'can' — in the relevant sense — do, it seems that the value primitivist has succeeded in securing access to a new normative term that applies to acts but not to eye colours, as desired.

I have two main objections to this. Firstly, it is not at all clear that this newly constructed concept really allows the value primitivist to make claims with any new *normative* impact. It's true that they can make a new claim that is partly normative, e.g. 'ϕ-ing is the right thing to do.' But this is decomposable into the conjunction of two claims: (i) the agent can ϕ, and (ii) no other option produces more value than ϕ-ing. Note that only the first claim is new (given a context in which we've already settled the value facts), whereas only the second claim is normative. So this move does not really allow the value primitivist to make additional normative claims in any interesting sense. It merely allows them the cheap victory of making conjunctive claims that are new in one respect and normative in another.

[12] I owe this suggestion to Nic Southwood.

My second worry about this proposal is that it seems *ad hoc*, or lacking a principled basis for being restricted in scope to actions. After all, the normative part of the new conjunctive claim — i.e., the conjunct which claims that the object maximizes value — is just as applicable to eye colours as it is to actions. Presumably the underlying idea was meant to be that the conjunction of 'value' and 'can' claims is normatively significant in a way that goes beyond the mere sum of its parts (so to speak). This is a plausible claim, but I do not think it is one that the value primitivist can make. If we have an *independent* grasp of rightness, then we can claim that *acts that are the best an agent can do possess this further normative property of rightness*. But such a claim would be trivialized by the value primitivist's conceptual reduction of 'rightness' to 'the best an agent can do'. It would be what Parfit (2011, 71) calls a 'concealed tautology': equivalent to the claim that *acts that are the best an agent can do possess the property of being the best that agent can do*. This is not yet to say that there's anything normatively significant about being the best an agent can do, over and above the evaluative component of the claim. So I don't think that value primitivists can escape commitment to the parity thesis so easily.

Even so, the value primitivist might at least cushion the blow by pointing out some pragmatic reasons why we might tend to be more *interested* in evaluating actions than, say, eye colours.[13] Actions are the means by which

[13] Thanks to Philip Pettit, Andrew Reisner, and Peter Singer for pressing me on this point.

we may voluntarily affect the world (and the amount of value in it). Moreover, our future behaviour may itself be influenced by praise and blame, whereas eye colours tend not to be so responsive. So even upon accepting the parity thesis, the value primitivist needn't face any problems explaining the act-centric focus of our ordinary moral practices. I'm happy to grant that. My objections to the parity thesis will emerge shortly (and more strongly in section 6, where I argue that we need an independent concept of rightness in order to maintain the consequentialism-deontology distinction). For now, I just want to re-iterate that value primitivists are committed to seeing no deep or principled difference in the range of normative assessments we can make of acts and eye colours, respectively.

The Fittingness view provides us with more options. As we have seen, value claims can be understood as claims about what it is fitting to desire. Now, it's true that *this* sort of normative assessment applies without restriction: we can ask whether some eye colour is fit to be preferred over others, just as we can ask whether some choice or action is fit to be preferred over others. But that isn't all that there is to say about what's fitting. In addition to these questions of desirability, we can also ask what it is fitting to believe, feel, and *choose*.

This last option gives rise to a form of normative assessment that applies distinctively to acts (understood as the direct implementation of a choice), as a matter of principle. That is, in addition to asking whether some chosen

act is fit *to desire*, we can ask whether it is *itself* fitting (qua choice).[14] No such additional question arises for eye colours. We can ask whether some eye colour is fit to desire, but there's no further question whether the eye colour is itself rationally fitting — eye colours just aren't subject to direct rational criticism in this way. It cannot be a rational failing for one's eyes to *be* a certain colour, the way it might be a rational failing (in some circumstances) to *choose to bring about* such an outcome.[15] The Fittingness view thus expands the scope of our moral theorizing to include a normative status distinctive to acts: whether they are fit to choose, or — in more colloquial terms — morally *right*.

I now want to argue that we should reject the Global Consequentialist's parity thesis, and hence favour the fittingness view over value primitivism. Consider again the completely general Global Consequentialist formula that *the 'right' x is simply the best (most utility-maximizing) x of those available*, where 'x' might be replaced by 'act', 'eye colour', or any other category of evaluand (Pettit and Smith 2000). I've been arguing that the Fittingness

[14] To clarify the distinction between choice and desire: I take desire to be a persisting mental state that has some associated degree of weight. Any one desire, insofar as it may be outweighed by others, leaves open the question of what to do. Choice, on the other hand, is the process of *settling what to do*. In the case of Buridan's ass, caught between two indiscernible bales of hay, I take it that the fitting response is to desire *that you have either*, but to choose either particular one. See section 6 for further discussion.

[15] One may question whether acts are really subject to direct rational criticism, given that their rational status would seem to derive from the rationality of the *choice* to so act. I'm happy to grant that choice is the more fundamental object of assessment. But insofar as it is natural to understand the choice as partly constituting the action, an assessment of the choice is *ipso facto* an assessment of the action. Choices and eye colours, by contrast, stand in no such intimate relation.

view shows there to be an important difference in normative structure here: there are some kinds of normative assessment that can be made of acts that cannot be made of eye colours. Though one may stipulatively define a sense of 'right' (or 'ought') that applies to everything including eye colours, there is an importantly natural, non-stipulative sense that applies only to judgment-sensitive evaluands (with no corresponding normative concept that applies just to non-judgment-sensitive evaluands). But this analysis will not convince someone who is not inclined to accept the fittingness view to begin with.

A more neutral way to bring out the difference in normative *structure* between actions and *mere* evaluands (like eye colour) is to think about how the relevant set of 'available options' is fixed in the above formula. In case of actions, the 'ought implies can' principle invokes a very particular sense of possibility: something like *rational power*, perhaps. That is, the relevant options are naturally restricted to something like *what the agent would succeed in doing if they had the intention to do it*[16] — though the precise details don't matter for my purposes.

The assessment of eye colours presents us with two notable points of contrast. First, there does not seem to be any naturally privileged set of options for mere evaluands; the relevant sense of 'availability' will instead presumably vary depending on conversational context. (For example, one speaker may be interested in the eye colour that one 'ought' to have amongst those genetically possible for a child of one's parents, whereas another may

[16] Thanks to Michael Smith for this suggestion.

be concerned to evaluate amongst all of those humanly possible. There's no fact of the matter as to which of these is the eye colour that you really 'ought' to have.) This looks an awful lot like ordinary evaluation — unlike in the case of actions, where the shift to 'ought'-talk brought with it a principled determination of the available options, and hence an unequivocal answer to the question of what act you ought to perform.

Secondly, eye colours are not judgment-sensitive, or directly subject to rational influence, the way that actions are. Judging that one ought to have blue eyes will not even tend to make it so. In this sense one cannot possess one's eye colour *for* (normative) reasons, the way that one can believe, desire, and act for reasons. We may demand that others justify their beliefs, or their actions, with adequate reasons. It does not make such sense to hold people (directly) accountable or answerable for their eye colours.

So while one might insist that there is some stipulative sense in which we "have reasons" to have the best available eye colour,[17] this verbal victory for the global consequentialist misses the point that there is a real difference in normative structure here that needs to be acknowledged. While we can evaluate anything, against any arbitrary set of 'alternatives', there is a special kind of normative assessment of rationally available actions that has no analogue for other (mere) evaluands. This difference allows us to make substantive claims about how we ought to act that are not just disguised evaluative claims. The same cannot be said for 'oughts' of eye-colour

[17] Thanks to Toby Ord and Michael Smith for pressing me on this point.

possession.

These facts suggest that in order to track the natural 'structure' of the normative domain — as per our third desideratum from §1.1 — normative theorists need a conceptual framework that offers asymmetric treatments of judgment-sensitive and non-judgment-sensitive evaluands. The Fittingness view achieves this by positing norms of 'fittingness' that apply directly to judgment-sensitive evaluands (belief, desire, action), one instance of which is the more specific notion of *fittingness to desire*, which in turn applies globally to all evaluands without restriction. The Fittingness view thus represents the normative domain as appropriately (asymmetrically) structured: our normative primitive of *fittingness* applies immediately to judgment-sensitive evaluands, and only indirectly (as mediated through the fittingness norms *for desire*) to everything else.

One needn't accept the Fittingness view in order to secure this desideratum, of course — a pluralist might combine the Moorean's value primitivism with an additional primitive concept of 'rightness'. But the Fittingness view is both more parsimonious and more powerful. We have seen that it is more parsimonious in unifying both evaluative and non-evaluative normative claims under the umbrella of the single normative primitive of 'fittingness'. What's more, it is more powerful in that it provides us with an explanation of *why* there is a distinctive form of normative assessment that applies only to acts. The explanation is that there are distinct forms of normative (fittingness) assessments corresponding to *each* of our judgment-sensitive rational

outputs: e.g., beliefs, desires, and actions.

The Fittingness view tells us that the scope of normative theorizing is fixed by the range of our rational outputs, as those are the things that can be assessed as more or less fitting.[18] At a minimum, we can identify assessments of *fitting belief* as the province of epistemologists, *fitting desire* as the concern of axiologists (value theorists), and — I suggest — *fitting choice/action* as a further issue for normative ethicists. This list should hopefully provide a plausible enough starting point, but I should stress that it is ultimately up to our best moral psychology / theory of agency to identify any additional rational outputs (e.g. emotions), and hence to further delineate the scope for normative inquiry.

4 'Privileging' Acts

Act Consequentialism is the view that the rightness of an action is determined by the value of the resulting state of affairs (compared to the available alternatives). We may further specify that, according to act consequentialists, acts are the *only* items to which this normative status of 'rightness' may apply.[19] There is, in this sense, an important normative disparity claimed between acts and (say) eye colours. We are now in a position to put to work

[18] Darwall (2006) offers a similar suggestion: "In principle, there are as many normative notions as items (action and attitudes) that can be normatively regulated."

[19] Technically, there may be an analogous status of 'rightness' or fittingness for other judgment-sensitive evaluands such as belief, etc. But epistemological questions and the like are outside the province of the moral theorist. The crucial point for now is just that the Act Consequentialist is naturally taken to reject the parity thesis, and to instead see an important asymmetry at least between acts and non-judgment-sensitive evaluands.

the theoretical apparatus developed in previous sections, illuminating the debate between Act and Global Consequentialists. In particular, we shall assess the charge that act consequentialism "privileges" actions in a theoretically objectionable way.

Let us first observe that *all* consequentialists start with an axiology, or theory of value, which tells us what things are valuable or fitting to desire. Given this starting point, the work that remains for their normative theory is to use these evaluative facts to derive some further normative claims about what we ought to do. We've seen that value primitivists are naturally led to the view that there isn't really anything more to say once we settle the value facts. All the substantive work is done by their axiology, and so to call an act 'right' is not to make a further claim at all, but is instead analytically equivalent to calling it 'best' or value-maximizing. The substantive content of this 'Global Consequentialism', over and above the axiology that might be shared by any other form of consequentialism, is purely negative: it's just the claim that there isn't anything more to say, beyond evaluating things (including acts) for their desirability.

But we've seen that there *is* more for a moral theory to say than just this. According to the fittingness view, in particular, we can assess acts not just for their desirability (whether they are fit *to desire*), but also for whether they *themselves* constitute fitting choices. This naturally supports Act Consequentialism over Global Consequentialism (modulo the qualms about 'rightness'-talk mentioned in footnote 4). Act Consequentialists begin

with an axiology (an account of fitting desire), and to this they add just one new type of claim — a claim about what it is fitting (right) to *do*. The claim "an act is right if and only if (and because) it produces at least as much value as any available alternative" is no longer a mere analytic truth or stipulative definition, but a substantive normative claim that relates one normative property (fitting choice) to another (fitting desire).

Act Consequentialism invokes a normative status that applies distinctively to acts, and thus succeeds in going beyond its axiology to make substantive new claims, when Global Consequentialism does not. This suggests two reasons to prefer Act Consequentialism: first, because it has more substance, and secondly, because it achieves this by better respecting the structure and scope for normative theorizing revealed by the fittingness view and our best moral psychology. When Global Consequentialists deny that acts are subject to a distinctive form of normative evaluation (beyond the generally available assessment of global desirability), this is both factually incorrect and unnecessarily limiting.

I've argued that Global Consequentialists have erroneously neglected the structural differences between normatively-assessable acts and mere evaluands. It's a mistake for a normative theory to treat everything on a par, because it turns out that there's a distinctive normative question that arises for acts — namely, 'are they morally right, or fitting to choose?' — that has no analogue in case of mere evaluands. But let me add a conciliatory note. Global Consequentialists are right to insist that our *axiological* theory

should assess each particular evaluand (including acts considered as mere evaluands) directly and in its own right. Many philosophers have made the contrary mistake of assuming that the best dispositions, for example, are simply those that result in the best actions.[20] To privilege actions in *this* (axiological) sense is indeed a straightforward error. It is important not to overlook the fact that dispositions of character can have good or bad effects other than via their manifestation in action (Adams 1976; Parfit 1984; Railton 1984; Pettit and Smith 2000). For example, it might be good, for the sake of deterrence, to have a transparent disposition to trigger a doomsday device when attacked. This does not entail that there's anything good about the act of mass destruction. The value here accrues from (transparently) possessing the disposition, not from exercising it.[21] So it would be a mistake to think that acts resulting from beneficial dispositions must themselves be beneficial. Each evaluand must be evaluated in its own right, as Global Consequentialists rightly remind us.

But just because actions should not be *axiologically* privileged, this does

[20] For example, Shaw (2006, 13) writes: "Consequentialists generally assess dispositions, behavioral patterns, and character traits in the same instrumental way: one determines which ones are good, and how good they are, by looking at the actions they lead to."

Even Kagan (2000) arguably suffers from this oversight. After introducing the idea of evaluating rules according to their propensity to produce more or less good actions, Kagan claims that this is tantamount to "claim[ing] that rules should be evaluated *directly* in terms of the goodness of their consequences." (p.149) But this only follows if the downstream actions are the only possible consequences of internalizing a rule, which we'll see is false.

[21] Geoff Brennan pointed out to me that this is still a case where the disposition's value derives from its effect on (others') *actions*. But we can just as well construct cases where, e.g., the mere possession of a certain disposition tends to make one happier, or to produce good effects in the world, in a way that is unmediated by any actions.

not mean that there are no important normative-theoretical differences between acts and (say) eye colours. Act Consequentialists can — and should — take care to avoid the axiological oversight that global consequentialists warn against, whilst maintaining that there's an important sense in which we can assess acts, but not eye colours, as morally right or wrong (over and above their being good or bad).

To summarize my argument thus far: The fittingness view implies that the list of rational outputs (states or activities that are responsive to our normative judgments, and can be assessed as more or less rationally 'fitting') determines the scope of our normative theorizing. There are facts about which features of an agent are rational outputs, and hence what kinds of reasons there can be (e.g. reasons for belief, desire, and action). These moral psychological facts are arguably 'prior' to the question of what particular normative theory correctly specifies the contents of these various kinds of reasons. So our normative theorizing should acknowledge and respect these limitations. Consequentialists presuppose an account of our reasons for desire (i.e., their axiology). They have no ambition to override epistemologists' claims about our reasons for belief, nor do they tend to make claims about the rationality of other attitudinal states like emotions. This would seem to leave reasons for action as the only remaining target for further theorizing. So it is by no means arbitrary for Act Consequentialists to add further normative claims (i.e. besides what is already contained in their axiology) *only* about how we ought to act. This is the only kind of claim that *can* plausibly

be added, once we have settled all questions about what we have reason to desire.

5 Analysing Rule Consequentialism

My analysis so far has presupposed that our theory of value settles all questions about what we have reason to desire: i.e., we should always prefer what's best.[22] However, deontologists and rule consequentialists might seem to deny this. (Here I'll focus on the latter for simplicity.) Rule consequentialists hold that we ought to act in accordance with certain rules (the general acceptance of which would maximize value) even on occasions when so acting is not itself value-maximizing (Hooker 2000). This seems difficult to make sense of if the value facts exhaustively specify our reasons for desire, since by calling one act value-maximizing (worth preferring) but then prescribing another as worth choosing, rule consequentialism would seem to imply that we ought *to hope* that we act differently from how we ought *to act*. Such a disconnect between rational preference and rational choice does not seem especially coherent (Portmore 2007, 50). Yet rule consequentialism is surely a coherent (if mistaken) view. What has gone wrong?

Most naturally, when the rule consequentialist prohibits the 'best' or 'value-maximizing' act, they do not really mean that the prohibited act is desirable all things considered, but only *antecedently* desirable, i.e. before we

[22] Note that our theory of value need not be impartial ('agent-neutral'). If we each have agent-relative reasons to prefer the welfare of our own children over that of strangers, for example, then different states of affairs will be desirable relative to different people.

consider the distinctive reasons for desire that derive from an act's deontic status as morally right or wrong. In this sense, their initial value theory is inconclusive or incomplete.[23] It accounts for only *some* of our reasons for desire: agent-neutral welfarist reasons, perhaps. But these reasons for desire are not decisive. Let's unpack how this might work.

Rule consequentialists first identify the rules that are best in terms of impartial welfare (or what's antecedently desirable), and then specify that we have decisive reasons to act in accordance with these rules. Finally, they might add, we have overriding reasons to desire that we so act. This way, a prohibited act may be 'best' according to the antecedent (agent-neutral welfarist) reasons for desire, and yet be bad (undesirable) all things considered. This avoids the incoherence mentioned above. But it also brings out how convoluted the view really is. It is recognizably consequentialist in the sense that it takes (*some*) reasons for desire as fundamental, and subsequently derives an account of reasons for action. But then it goes back and "fills in" further reasons for desire — trumping the original axiology — to make sure that they fit the account of right action. In this sense it exhibits a deontological streak: reasons for action are at least *partly* prior to reasons for desire. In other words, the initial axiology includes only some values (the 'non-moral', agent-neutral welfarist ones), and what's *right* serves to determine the remaining ('post-moral', all things considered) good.

On this analysis, the distinction between deontological and consequential-

[23] Thanks to Michael Smith for suggesting this interpretation.

ist moral theories is a matter of the relative priority they assign to reasons for action and reasons for desire. This is an updated version of the traditional idea that deontologists take 'the right' as prior to (and partly determinative of) 'the good', whereas consequentialists think we can explain what's right entirely in terms of promoting non-moral goodness. Though the right and the good are clearly connected in some way, we may wonder which is more fundamental: which explains, or is the basis for, the other.

One appealing aspect of this analysis is that it allows us to reject recent arguments from the possibility of agent-relative value to the conclusion that really *all* theories are consequentialist (cf. Louise 2004). Critics of the distinction have noted, for example, that a deontological prohibition on lying may prove extensionally equivalent to a consequentialist injunction to maximize the agent- and time-relative value of *being yourself now honest*. But we are not thereby forced to collapse the distinction between deontology and consequentialism, so long as we can distinguish the two possible ways the relative priority might work out. The consequentialist version of the view would say that one has reason to act honestly *because* of the antecedent desirability of being oneself honest. A deontologist, by contrast, might claim the reverse: that the desirability of being oneself honest instead derives from the independent rightness or choiceworthiness of such acts. I think it is important to maintain this distinction. For while some agent-relative views, e.g. ethical egoism, may be naturally understood as goal-directed or consequentialist in nature, it seems a distortion to suggest that traditional deontological views

are really just about promoting the goal of one's own moral purity. (Even if that accurately describes their *upshot*, it does not seem not a fair characterization of what deontological theories *present* as morally significant.)

6 Does Fitting Choice Collapse into Desirability?

We are now in a position to address what I see as the most pressing objection to the argument of this paper. I've been arguing that the fittingness view allows us to expand the expressive power of our normative theorizing, by raising new questions of 'fitting choice' over and above questions of value or desirability. But one might object: *why think that there are further normative facts about 'fitting choice', over and above the facts about fitting desire?*[24] To bring out the problem, let us imagine that an agent is faced with a choice between two actions: she may either ϕ or ψ. Further suppose that she has determined that ϕ-ing is the option that is all things considered most desirable. Is there any further question here of what she should choose? It may seem not. Again, it would seem somehow incoherent to think that ϕ-ing is the most desirable action, and yet that she ought to ψ instead. Choosing to ψ is clearly in tension with wanting, all things considered, to ϕ; it seems that one or other attitude must be mistaken (inappropriate, unfitting). This may be taken to suggest that questions of fitting choice are reducible to questions of fitting desire, so that there is not really any 'further fact' here of the sort that I have suggested.

[24] Thanks to Michael Smith for pressing me on this.

I have three responses to this objection. Firstly, as discussed in footnote 14, there may be cases like 'Buridan's ass' in which an agent must choose between equally desirable options. In such a case, it seems that either choice (of one over the other) would be fitting, though it would not be fitting to prefer either option over the other. This suggests a subtle difference between the norms that govern preference and those that govern choice, albeit one that may only show itself in a very limited class of situations.

Secondly, even in the remaining cases, I do not think that it is strictly *self-contradictory* to claim that what it is fitting to choose diverges from what it is desirable to do. Instead, like the claim *that suffering is good*, or *that only the welfare of a certain race of people matters*, I take it to merely be a very obvious but non-analytic falsehood. One could hold the contrary view without contradiction; it just isn't remotely plausible as a matter of substantive normative fact. So our imagined agent, having determined that it's most desirable that she ϕ, does in fact face a further question of whether she should choose to ϕ. It merely happens to be a very easy question.

Thirdly, even if one thought that fitting choice and fitting desire must coincide as a matter of conceptual necessity, no conceptual *reduction* is possible if the direction of explanation remains a conceptually open question, as I think it clearly does. Although act consequentialists think that acts are choiceworthy in virtue of being all things considered desirable, my previous section argued that we should understand rule consequentialists and deontologists as reversing the order of explanation. In the case of rule con-

sequentialists, we find that some things (e.g. aggregate welfare) are indeed antecedently desirable, but then facts about what's choiceworthy intervene to influence what is ultimately all-things-considered desirable. Deontologists might even go so far as to take the facts about choiceworthiness as wholly fundamental and determinative of what's desirable. Now, I take it to be a matter of substantive normative fact which of these is the true moral theory. It would be prejudicial for our conceptual framework to take a stand on this issue. So we should acknowledge that there is room in our normative theorizing for both 'fitting choice' and 'fitting desire' assessments, and that it is a matter of substantive normative fact which of these is prior. There are two distinct conceptual possibilities here, as there would not be if the distinction between fitting choice and desirability of choice were to collapse. So the distinction must be upheld.

One might object that my earlier arguments against Global Consequentialism show that the Fittingness view itself violates the desideratum of neutrality between first-order normative theories.[25] But I think my analysis is better understood as showing that Global Consequentialism is not a purely first-order normative view (in the relevant sense). As we saw in section 4, we can decompose Global Consequentialism into two components: a first-order axiology that might be shared by any other form of consequentialism, and the structural claim ('structural' in the sense found in our third desideratum from §1.1) that *the normative domain is flat and symmetrical in structure,*

[25] Thanks to an anonymous referee for suggesting this objection.

such that judgment-sensitive and non-judgment-sensitive evaluands are on a par. It is only this latter claim that I object to.[26]

Recall the motivation for neutrality: we wish to select a conceptual framework within which to conduct our first-order inquiry. Even if it were a count against the Fittingness view that it rules out Global Consequentialism, that's clearly still a huge improvement over the value primitivist's exclusion of *every* form of non-consequentialism. But given my analysis of Global Consequentialism in this paper, I don't think we should really see it as a cost at all. We've seen that, so far as its first-order normative claims are concerned, Global Consequentialism does not make any further claims beyond Act Consequentialism.[27] And the fittingness view is neutral with respect to *those* claims. It is only the Global Consequentialist's assertion of the parity thesis that is rejected by the Fittingness view. But that's not a problem, because one of the desiderata for our conceptual framework is to settle this 'structural' question about the natural joints of the normative domain. We did not *want* to remain neutral on that.

Another way to make the point is that adopting value primitivism as one's conceptual framework effectively just begs the question against non-

[26] And even this is not strictly incompatible with the Fittingness view alone. As explained in section 3, the Fittingness view provides a formula for determining the scope of normativity: there will be a kind of normative status corresponding to each kind of rational output. So the result that there are other normative assessments to be made besides assessments of desirability only follows from the Fittingness view once it is supplemented by the further (albeit obvious) claim that there are other rational outputs besides desires.

[27] If anything, it may — as we've seen — make fewer claims, if in calling an act 'right' they are *merely* saying that it maximizes value, and not further claiming that it thereby has the distinct property of being fitting to choose.

consequentialist views. My arguments for the Fittingness view, by contrast, have (if successful) *shown* why the structural elements of Global Consequentialism are mistaken. I have not simply assumed it. So I do not think the Fittingness view is objectionably prejudicial in this regard.

Conclusion

In this paper, I have argued that we should take *fittingness*, rather than *value*, as our normative primitive. This allows us to say everything that we could have said using value-talk, and more besides. It helps us to avoid the misguided global consequentialist idea that judgment-sensitive evaluands (e.g. choice or action) are 'on a par with' — or normatively assessable in all the same ways as — mere evaluands such as eye colours. Instead, we find that there's a distinct property of rightness that applies only to acts, over and above the property of value (or desirability) that all sorts of evaluands might have. Finally, the fittingness view, unlike value primitivism, provides an appropriately ecumenical conceptual framework that remains neutral on the first-order normative dispute between consequentialism and deontology.

I further argued that the fittingness view is preferable to the pluralist view that takes both *value* and *rightness* as independent primitives. This is because the fittingness view is more parsimonious, capturing both kinds of assessment under a single umbrella. And it is more explanatory, in that it provides us with an account of *why* there is a distinct normative status corresponding to acts (namely, because there is a distinct normative status

corresponding to each class of judgment-sensitive rational outputs, and because actions, alongside beliefs and desires, are one such type of output). Overall, the fittingness view provides an elegant conceptual framework that offers everything we want from a normative primitive.

References

Adams, Robert Merrihew. 1976. "Motive utilitarianism." *Journal of Philosophy* 73:467–481.

Brentano, Franz Clemens. 1902. *The Origin of the Knowledge of Right and Wrong*. Westminster, A. Constable & co, ltd. Translation of Vom Ursprung sittlicher Erkenntnis. English translation by Cecil Hague.

Darwall, Stephen. 2006. "How Should Ethics Relate to (the Rest of) Philosophy? Moore's Legacy." In Terry Horgan and Mark Timmons (eds.), *Metaethics after Moore*. Oxford: Oxford University Press.

Dreier, James (ed.). 2006. *Contemporary Debates in Moral Theory*. Malden, MA: Blackwell.

Hooker, Brad. 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Oxford University Press.

Hooker, Brad, Mason, Elinor, and Miller, Dale (eds.). 2000. *Morality, Rules, and Consequences*. Edinburgh: Edinburgh University Press.

Kagan, Shelly. 2000. "Evaluative Focal Points." In Hooker et al. (2000).

Lawlor, Rob. 2009. *Shades of Goodness: Gradability, Demandingness and the Structure of Moral Theories*. Palgrave Macmillan.

Louise, Jennie. 2004. "Relativity of value and the consequentialist umbrella." *Philosophical Quarterly* 54:518–536.

Moore, G.E. 1903a. *Principia Ethica*. Cambridge: Cambridge University Press.

—. 1903b. "Review: The Origin of the Knowledge of Right and Wrong." *International Journal of Ethics* 14:115–123.

Norcross, Alastair. 2006. "Reasons Without Demands: Rethinking Rightness." In Dreier (2006).

Ord, Toby. 2009. *Beyond Action: Applying consequentialism to decision making and motivation*. Ph.D. thesis, University of Oxford.

—. ms. "How to be a consequentialist about everything."

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

—. 2011. *On What Matters*, volume 1. Oxford: Oxford University Press.

Pettit, Philip and Smith, Michael. 2000. "Global Consequentialism." In Hooker et al. (2000).

Portmore, Douglas W. 2007. "Consequentializing moral theories." *Pacific Philosophical Quarterly* 88:39–73.

Rabinowicz, Wlodek and Rønnow-Rasmussen, Toni. 2004. "The strike of the demon: On fitting pro-attitudes and value." *Ethics* 114:391–423.

Railton, Peter. 1984. "Alienation, consequentialism, and the demands of morality." *Philosophy and Public Affairs* 13:134–171.

Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Belknap Press of Harvard University Press.

—. 2009. "Being Realistic about Reasons." John Locke Lectures.

Shaw, William. 2006. "The Consequentialist Perspective." In Dreier (2006).

Shope, Robert K. 1978. "The Conditional Fallacy in Contemporary Philosophy." *The Journal of Philosophy* 75:pp. 397–413.

Sidgwick, Henry. 1907. *The Methods of Ethics*. Thoemmes Press.

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

—. 2010. "Beyond the Error Theory." In Richard Joyce and Simon Kirchin (eds.), *A World Without Values: Essays on John Mackie's Moral Error Theory*, 119–139. New York: Springer.