On a Theory of a Better Moral Theory and a Better Theory of Morality[1]
by
Geoff Sayre-McCord
(draft of May 2018)

Introduction

Normally, there is a sharp distinction between a better theory of X and a theory of a better X (and so between the best theory of X and the (best) theory of the best X). That it would be better if everyone were treated as equals is no reason whatsoever to think that they are; that it would be better that death came quickly, painlessly, and late in life is no reason whatsoever to think it does; that it would be better if we could fly with the snap of a finger is no reason whatsoever to think that we can… That the theory of a better X is a theory according to which things are different from the way one's (so far) best theory says they are is (normally) no reason whatsoever to think one's (so far) best theory is wrong, just reason to wish X were different (and, if it is possible, reason to work to change X).

In contrast (I claim) when the subject matter (the proverbial X) is normative, this normally sharp distinction is elided. There being a difference between one's theory of the best normative X (the best morality, the best standards of inference, the best rules of justification…) and one's (so far) best theory of X, necessarily provides a reason (though perhaps not a decisive reason) to think one's (so far) best theory is wrong. The elision, I suspect, plays an essential role in a range of arguments concerning morality, practical rationality, and theoretical rationality, a few of which I discuss below.

Yet the elision smacks of depending crucially and unacceptably on wishful thinking – on supposing that the fact that things would be better if only they were a certain way provides some reason to think they are that way. Does the elision involve an unacceptable version of "Wouldn't it be nice that p, therefore p" reasoning? On the contrary, I argue, it is as a reflection a constraint on acceptable normative theories that is itself explained by a distinctive characteristic of normative concepts, one that sets them all apart from descriptive concepts.[2]

As a first step in making that argument, though, let me turn, with more detail, to the elision to which I have alluded.

[2] Here I am using the phrase "normative concept" as a term of art that is vindicated not by common usage but -- if at all -- by it capturing a theoretically valuable contrast. I come back to whether it does in the conclusion of this paper.

A Normally Sharp Distinction

Normally, as I have said, there is a sharp distinction between a theory being a better theory of X and it being a theory of a better X. Suppose, for instance, that we develop a theory of what the laws of (say) North Carolina are. This is of course no easy task. We would need to take account not just of the laws on the books, but whether and how they were enforced, as well as the nature and grounds of past judicial decisions (and the weight of *stare decisis*), the relevance of common law and the distinction between that and statutory and regulatory law, of which "violations" were actually punished, and which just "taxed," of whether various candidate laws conformed to the standards of legal legitimacy in force in the state, of how state law is related to federal law, etc..[3] Imagine, though, that we (tentatively, but with some confidence) arrive at a theory of what the laws actually are.

Next, suppose we turn to the project of figuring out what the best set of laws for North Carolina might be. Again, this is of course no easy task. Empirical and normative complexities abound. We will need to take account, presumably, of whether and how easily the laws might be understood, of what enforcing them would involve, of their compatibility with Federal laws, of their capacity to win willing compliance, of the effects of putting them in place, of the degree and nature of their infringement on liberty, of whether they are fair and fairly enforceable, etc.[4] These complexities notwithstanding, imagine that we (tentatively, but with some confidence) arrive at a theory of what the best laws would be.

Imagine too that our two theories do not coincide perfectly – that on our understanding of what the laws actually are, the laws of North Carolina would be better if only some of them were different. (Not an implausible conclusion.) Would this give us any reason to think that our original theory of what the laws are is mistaken. Normally not[5]; rather, it would give us reason to wish the laws were different and reason to work to change the laws (assuming they can be changed).

The same is true if we were to switch from a theory of the laws of North Carolina to one concerning the workings of the mind, or the human body, or social conventions. In each case, our coming to think things would be better were they different than we take them to be is, it seems clear, *no reason at all* to think that they are different – only reason to wish they were different and to change them if we are able. There is, in other words, a sharp distinction between a theory of a better mind, of a better human body, and of better social conventions, on the one hand, and a better theory of the mind, of human bodies, and of

---

[3] Here I mean to leave room for whatever sorts of considerations people think bear on a proper understanding of what the laws *actually are.*

[4] Here I mean to leave room for whatever sorts of considerations people think bear on a proper understanding of what the laws *should be.*

[5] The "normally" is important here, since if our theory of the nature of North Carolina's laws includes the claim that certain normative requirements are a part of the law the shortfall in the law as we take it to be may actually serve as grounds for thinking we have misunderstood what the law actually is. More about this in what follows. For now, assume that no normative elements appear in our positive theory of the laws of North Carolina, or that the ones that do are not elements that tell in favor of the laws we have identified as better.

social conventions, on the other. The first tells us how it would be good for things to be, the second how things are.

The Distinction Elided

So much is true, *if* the theories are of non-normative matters, but not, I believe, otherwise. Suppose, for instance, that we develop a theory of what the standards of morality are. This is of course no easy task. We would need to take account not just of what is intuitively plausible, but (presumably, though this is somewhat contentious) of what might be brought under general plausible principles that are mutually consistent, of what is metaphysically tractable, of what is epistemically accessible, etc.[6] Imagine, though, that we (tentatively, but with some confidence) arrive at a theory of what the standards of morality actually are.

Next, suppose we turn to the project of figuring out what the best standards of morality might be.[7] Again, this is of course no easy task. Complexities abound: at least all those already mentioned concerning our attempt to develop a theory of morality.[8] Still, imagine that we (tentatively, but with some confidence) arrive at a theory of what the best standards of morality would be (by the lights of our original theory). Imagine too that our two theories do not coincide perfectly – that on our understanding of what morality's standards actually are, a different set of standards would be (morally) better.[9] Would this give us any reason to think that our original theory of what the standards of morality are is mistaken. I think it would; we would be finding ourselves with reason to think we had originally gotten the standards wrong, not with reason to wish the standards of morality were different and reason to work to change those standards (if only that were possible). In this case, we would have found there was reason to guide our judgments, actions, and reactions by the morally better standards, rather than what we had originally taken to be the standards of morality.

---

[6] Here I mean to leave room for whatever sorts of considerations people think bear on a proper understanding of what the principles of morality actually are.

[7] The quality of a moral or legal system is a direct reflection, I am assuming, of the quality of the standards or laws that comprise it. But, I am also assuming, the quality of those standards or laws may well depend in part on how they play out in practice – say, on whether the moral standards can secure allegiance and whether the laws of a legal system can be consistently enforced. To what extent how things would play out in practice matters to the evaluation of the standards or laws in question is of course a contentious issue. For my purposes, it is to be settled by our best theory of morality or of the law, respectively, since the relevance of the theory of a better X to the best theory of X depends on the (so far) best theory including the standards in light of which the other theory counts as better.

[8] If I am right that the project of developing a theory of the best morality is part and parcel of developing the best theory of morality, the complexities involved in doing the latter will involve all those involved in the former. Specifically, the burden of evaluating how the standards we arrive at fare, morally, in comparison with alternatives, which I highlight in discussing which morality is best, has a place in determining which theory of morality is the best.

[9] As I stress below it matters that the considerations in light of which one morality counts as better than another are themselves moral considerations in order for the elision to be induced. Similarly, in other normative domains, the relevant standards of betterness are those endemic to that domain.

Significantly, the legal example switches sides, so to speak, if (i) our best theory of law incorporates moral principles as central to what the laws are and (ii) it is in light of those principles that our theory of a better set of laws counts as better. This may happen, for instance, on Ronald Dworkin's theory of law according to which equal concern and respect are central to legal legitimacy in the United States. In that case, if our best theory of North Carolina's laws include a prohibition on same-sex marriage a commitment to equal concern and respect tells in favor of allowing such marriages we would thereby have reason (albeit not necessarily conclusive reason) to think our original theory might be mistaken about the prohibition on same sex marriage. Evidence against the prohibition is provided by the standards that, according to our best theory of the law, are a part of the law and serve as constraints on other laws.[10]

Some Examples Concerning Morality

To put a little flesh on the bones of what I have in mind, let me mention several examples of arguments that try to take advantage of the elision between a better theory of morality and a theory of a better morality to establish their conclusions. As will become clear, not all these arguments can be sound – since in some cases they are for incompatible conclusions – and none of them may be. What matters for our purposes is not whether the arguments go through but the form of argument they share.

One example can be found in Adam Smith's discussion of fortune's effect on praiseworthiness and blameworthiness. Smith starts out with a crisp and elegant argument against "moral fortune." Articulating a familiar thought, he argues that we are justly praised or blamed, approved or disapproved, only for what is under our control. And then he maintains that what is due to fortune (that is, to luck, good or bad), is not suitably under our control and so not properly attributable to us.[11]

Strikingly, though, having given this argument, Smith immediately turns to canvassing all the ways in which our praise and blame, approval and disapproval, are actually influenced by fortune. Whatever principles we might embrace in the abstract, in practice, he points out, we regularly take account of things that are not in peoples' control when praising and blaming, approving and disapproving.

Smith finishes his discussion of moral fortune with a robust defense of this practice. The defense plays out against the following observation

> If the hurtfulness of the design, if the malevolence of the affection, were
> alone the causes which excited our resentment, we should feel all the furies

---

[10] Similarly, if the (putatively) best theory of law is as Natural Law theorists claim, then in discovering that what we initially took to be the law was, in fact, not acceptable in light of Natural Law, we would have reason to think we had gotten the law wrong.

[11] This leads Smith to hold that "To the intention or affection of the heart… all praise or blame, all approbation or disapprobation of any kind, which can justly be bestowed upon any action, must ultimately belong." Only our intentions and affections of our hearts are, he thinks, suitably under our control. *The Theory of Moral Sentiments*. p. 133

of that passion against any person in whose breast we suspected or believed such designs or affections were harboured, though they had never broken out into any actions.[12]

And his defense turns crucially on the claim that "if the baseness of the thought which had given birth to no action, seemed in the eyes of the world as much to call aloud for vengeance as the baseness of the action, every court of judicature would become a real inquisition." Only by allocating praise and blame on the basis of actions and their effects, and not merely on an agent's (supposed) intensions and designs, can we, Smith argues, avoid such inquisitions and successfully protect privacy.

Smith never returns to his original argument against moral fortune nor to his claim that

> The consequences which actually, and in fact, happen to proceed from any action, are, if possible, still more indifferent either to praise or blame, than even the external movement of the body. As they depend, not upon the agent, but upon fortune, they cannot be the proper foundation for any sentiment, of which his character and conduct are the objects.[13]

So we never get a clear indication of what, in his original argument, Smith would reject. Nonetheless, Smith seems to think that the importance of privacy, of avoiding an inducement to inquisition, and of providing an incentive not to act on one's malevolent one's attitudes, all recommend standards for praise and blame that take actions and their effects into account (even as intentions too are considered).[14] In effect, having first set out a principled ban on moral fortune as a part of the best theory of morality, Smith offers apparently moral grounds for thinking a better morality would allow the influence of fortune in the allocation of praise and blame, from which he seems to infer that morality actually does allow it.

Appealing to entirely different grounds for the same conclusion, Margaret Urban Walker argues that leaving room for moral luck is morally important.[15] Specifically, she argues that acknowledging moral luck is the only way to accommodate the kind of person whose agency is ineluctably bound up with not just her own intentions but with (some of) the causes and effects of her actions, say of what her country has done or the effects of her actions that could not be anticipated. Such an agent properly takes responsibility, and can properly be held responsible, for more than what is directly under her control. Allowing that there are such agents involves, on Walker's understanding, rejecting the "control condition" on responsibility that seems so strongly recommended by our best theory of morality. Yet

---

[12] ibid. pp. 151-152.
[13] ibid., pp. 133-134.
[14] It is important to note that Smith is defending a focus on actions and their effects not just within our judicial system but across the board in our judgments of praise and blame. His reference to "every court of judicature" pretty clearly is meant to cover every context in which we judge of one another (and ourselves), and not restricted to the legal system.
[15] See her "Moral Luck and the Virtues of Impure Agency," *Metaphilosophy* 22 (1991), pp. 14-27.

we must reject it, she argues, since without recognizing this sort of extended agency there is no room for the notable moral virtues (of, for instance, dependability and integrity) that require its availability and exercise. In effect, then, having acknowledged that the "control condition" is understandably a central aspect of our best moral theory (at least until we see that there is a morally better alternative), Walker offers moral grounds for thinking a better morality would allow the influence of fortune in determining who is responsible for what. And she takes that as grounds for thinking that in fact there is moral luck.

Using the same form of argument, albeit in support of the opposite conclusion, people regularly argue that while our particular moral judgments might recommend as best a moral theory according to which there is moral luck, there is a moral objection – that in effect appeals to the control condition – to countenancing moral luck: that holding people responsible for what is beyond their control is unfair. They infer that this unfairness constitutes grounds for rejecting the theory our particular moral judgments would support. The crucial idea here is that a better morality – a fairer morality – is one that restricts an agent's responsibility to what she has control over.[16] And this is grounds, people take it, for thinking morality actually does restrict responsibility in this way.

Switching from moral luck to moral dilemmas, one might well raise moral objections to the idea that people are justifiably blamed for acting in situations in which they have no better option. Indeed, I have offered an argument along these lines, holding that for any dilemma-allowing moral theory there is a counterpart non-dilemma-allowing theory (albeit not the same other theory for each) that is morally better. They are morally better, I argue, because they avoid unfairly condemning people for failing to do the impossible. Moreover, assuming the dilemma-allowing theory acknowledges that one has a moral objection against a system, or institution, or person, sanctioning you for failing to do the impossible, the non-dilemma-allowing theory will be better than the original theory's own lights. And since every (plausible) dilemma-allowing theory is bettered by one that does not allow dilemmas, I argue, we have reason to think there are no dilemmas (even as there are wrenching choices that may well be between pressing but incommensurable demands).[17]

Another appeal to fairness shows up in Liam Murphy's views concerning whether and how our duties to help those in need change when others fail to help. According to him, considerations of fairness tell specifically against a theory according to which our duties increase directly in proportion as others fail to do their duty. His idea is that the duty of beneficence is a duty concerning which we each share in the responsibility for seeing it satisfied. In cases where someone is failing to do their share, a set of principles that required people to carry someone else's load would simply be unfair. It would be treating the moral shortfall of some as grounds for shifting the burden to others.[18] That this would be unfair is taken by Murphy as grounds for thinking it isn't so.

---

[16] Walker of course can acknowledge this cost while arguing that on balance making room for the possibility of impure agents more than compensates.
[17] In a "Moral Argument Against Moral Dilemmas," unpublished manuscript.
[18] See Moral Demands in Nonideal Theory.

Turning to a different issue, but sticking with the same kind of argument, Warren Quinn maintains that people have rights on the grounds that "it is fitting that they should."[19] In other words (I think): it is good (that is: fitting) that people have rights, therefore we have reason to think that they do. Or, to put it in still another way: a moral theory according to which people have rights is a theory of a better morality, a more fitting morality, so (he infers) a better theory of morality is one according to which people do have rights.

Inspired by Quinn, Thomas Nagel embraces this line of argument, to the same end, defending basic moral rights that reflect our status as inviolable on the grounds that a morality that acknowledges such rights is a better morality than one that doesn't. "The argument is that we would all be worse off if there were no rights… – ergo, there are rights."[20] According to him, this is because "one morality is more likely to be true than another, because the former makes for a better world than the latter — not instrumentally, but intrinsically."[21]

Finally, to take one last example, Michael Slote defends our right to pursue our personal aims and projects rather than, and at the expense of, the public good, on the grounds that the "moral autonomy" such a right allows is intrinsically valuable. Any theory that rejects such a right – act consequentialism, for example – offers standards that are morally worse than those of alternatives that recognize the right. So a theory of a better morality is a theory of a morality that affords people the right to pursue their private aims rather than the public good.[22] And from that he concludes that a better theory of morality is a theory according to which morality affords such a right.

Our interest here, of course, is not in the soundness of any of these particular arguments but in the argument form they share. Should we think that a difference between our (currently) best theory of morality and our theory of a better morality gives us reason to revisit our understanding of morality? If so, why?

Before turning to this, a few observations are in order.

The first is that such arguments are plausible at all, it seems, only if the dimension along which one (putative) morality is supposed to be better than another is recognizably moral. That one rather than another is financially, or self-interestedly, or legally, better is no reason to think that it is the actual one (unless, perhaps, these improvements are morally important).

Moreover, in order for the theoretical availability of a better morality to provide *us* reason to revisit our (current) best theory of morality the value at stake must be one that we

---

[19] Morality and Action, (Cambridge: Cambridge University Press, 1991), p. 173.
[20] This is true, he maintains, whether or not the rights are violated.
[21] "Personal Rights and Public Space," *Philosophy and Public Affairs* 24 (1995), p. 92. Nagel emphasizes that the value in play here is, specifically, not instrumental. He is not arguing that people having rights has, or will have, good consequences, but rather that their having rights is intrinsically valuable. See also Frances Kamm's Morality, Mortality: Volume II: Rights, Duties, and Status (Oxford: Oxford University Press, 1996), Chapter 10.
[22] See Common-Sense Morality and Consequentialism. (London: Routledge and Kegan Paul, 1985), Chapter 2.

can be brought to recognize. The most elegant case of this is when the moral value that is in play is already recognized as morally important by our (current) best theory.[23] But the arguments presumably get their force (assuming they have any) not from our *recognizing* that one (conception of) morality is morally better, but from it being so. And any justification of the argument should, as a result, trace back to a link between the actual standards of morality and the moral credentials they enjoy.

Finally, that we have a theory of a morality that is morally better than that represented by our (currently) best theory *in some respect* is compatible with that morality being morally worse in other respects. And those other respects might well outweigh (or undermine) the reason we have for thinking our (currently) best theory is wrong. When it comes to each of the arguments I have mentioned, their force depends not just on the moral value of what is on offer by the alternative theory, but on what, if anything, of moral significance is sacrificed by it.

Still, keeping these considerations in mind, it seems as if sometimes we might well find ourselves thinking that our (currently) best theory of morality does not, after all, align with our theory of the best morality. In these cases, I contend, we will have found a reason (albeit not necessarily a conclusive reason) to think our best theory has gotten morality wrong. What is going on?

There is really no question that part of what is going on is the pursuit of reflective equilibrium in action. That pursuit starts with a range of more or less specific moral judgments -- concerning, say, moral luck, or moral dilemmas, or the significance of rights, or the value of autonomy – and seeks out more general principles that can fit with and justify those judgments. Wherever the fit is poor or the justification weak there is reason to revise either our confidence in the judgments or the principles we advance as justifying them. This is a familiar process, one that reflects the idea that our particular judgments are *pro tanto* relevant evidence for or against various moral principles and that the ability to articulate principles that would justify our judgments is *pro tanto* relevant evidence for those judgments.

To the extent arguments are simply of the form 'here is something important that your theory does not accommodate,' this may be all that is going on. Interestingly, though, in these arguments what is being claimed is not simply that some value is being missed. In fact, in each case we can imagine that the (so far) best theory of morality acknowledges and provides an account of the values that are being appealed to in defending the alternative theory as a theory of a better morality. Rather, what seems to be going on is that the

---

[23] Smith's, Walker's, and my arguments all work with considerations – the importance of privacy, of integrity, and of fairness – that pretty much all plausible moral theories recognize as being of moral importance. Indeed, when it comes to my argument against moral dilemmas, many who defend dilemmas willingly acknowledge that they are unfair but argue that their unfairness is the essence of the tragedy they engender. In contrast, Quinn's, Nagel's, and Slote's arguments, which appeal to the fittingness or intrinsic value of rights, or the inviolability or autonomy they make possible, are likely to meet opposition from the start from those who embrace various influential versions of consequentialism, though not all. Even in these cases, of course, if these things are morally valuable then there will be reason, even if consequentialists do not recognize that reason, to judge the alternative theories better, at least in certain respects.

(acknowledged) *value of* a value, or a right, or a permission, is itself being offered as grounds for thinking the alternative morality is morally better. And it is this aspect of the arguments that invite the worry that they are engaging in wishful thinking and not introducing new reasons for their conclusions. Yet what is going on is, I think, something else and something important: attention is being shifted from the fit between the principles offered by the theories and more particular judgments, to the moral evaluation of those principles (where the fit between the principles and the more particular judgments is left unchallenged). What is at stake is more a matter of whether the original principles genuinely justify the more particular judgments, given that there are morally better alternative principles, not whether they fit with those judgments.

This is not to say the arguments works independently from a reliance on reflective equilibrium. I suspect that no arguments concerning morality, or other normative subjects, or scientific ones, for that matter, could do that. And, in any case , I do not think these arguments do. But within the gamut of reflective equilibrium there is a great wealth of distinctions to be drawn among arguments (between those that are deductive and those that are inductive), among deductive arguments (between those that rely on modus ponens, and those that rely, say, on conjunction elimination), as well as among arguments that are of the same kind and form (between the sources and the credentials for the various premises on which they rely). The question is whether there is a distinctive kind, or form, of argument in play in these cases, one that might render them plausible and worth taking seriously while perhaps also revealing something distinctive and interesting about normative theory.

<u>If it would be good were something good, does that show that it is good?</u>

One elegant suggestion, made by David Enoch, is that at least some moral terms express operators that respect a rule analogous to the following concerning necessity: that *if it is necessary that it is necessary that p, then it is necessary that p* ($\Box\Box p \to \Box p$).[24] The idea is that, for instance, "it is good that…" might be such that *if it is good that it is good that p, then it is good that p* ($GGp \to Gp$).[25] If so, then we could show the actual moral value of something by showing the moral value of it being morally valuable. Assuming the rule is in place for "it is good that…" we would have, if not quite an argument from wishful thinking, an argument of the form "Wouldn't it be morally good that *p*, for a moral *p* (specifically a moral *p* to the effect that something is morally good)," to "therefore *p* (that is, "therefore something is morally good.").

If only "it is good that *p*" entailed that *p*, in the way "it is necessary that p" entails p or "A knows that *p*" entails p – that is, if only value claims were factive – then $GGp \to Gp$ would follow straight-away; just as Gp would entail *p*, so too would GGp entail Gp. But normative claims, including those concerning what is good, or justified, or obligatory, are not

---

[24] "Wouldn't It Be Nice If *p*, Therefore, *p* (for a moral *p*)," *Utilitas*, vol. 21, no. 2, 2009, p. 222-224, p. 223.
[25] At least if we hold fixed that it is moral goodness that is in play.

in this sense factive. That *p* is good, or justified, or obligatory (or right or virtuous or …) is, sadly, perfectly compatible with ~*p*.

In this respect, at least, talk of something being good is analogous to talk of something being possible – no grounds at all for thinking what is good or possible is actual. So the question is whether there is some explanation of why GGp → Gp even though ~(Gp → p). As with possibility, concerning which it is similarly true that ~(◊p → p), some account is needed for why the relevant operator, iterated, entails its application.

The analogy with possibility, as Enoch notes, gives hope. For while *it is possible that p* does not entail *p*, on a very standard way of understanding possibility, that *it is possible that it is possible that p* does entail *it is possible that p*: (◊◊p → ◊p). The idea is that *it is possible that p* entails *p is true in some possible world* (whether or not it is true in the actual world). As a result, *it is possible that it is possible that p* entails that *it is possible that p is true in some possible world* (whether or not it is true in the actual world). But that entails *it is true that p is true in some possible world*, which is equivalent to *it is true that it is possible that p*, or, in other words, that *it is possible that p*. So, *it is possible that it is possible that p* entails that *it is possible that p*, even though thoughts about possibility in general are not factive.[26]

Yet when it comes to "it is good that…" and other moral operators, there is reason to doubt any such explanation is available, because there is reason to doubt the entailment. Consider the apparent fact that people are not only less than perfect, some people are not morally good at all. Thinking just of those people, it seems clear that it would be morally good if they were morally good. But that neither entails nor even gives us any reason to believe that they are morally good after all. So GGp (where Gp=the people in question being morally good) is true, though patently Gp is not. It seems as if there are two different ways Gp might have been true: if either the people were different than they are or the moral standing of how they actually are were different. That it would be good if one or the other were true seems to give us no reason at all to think either is. Other cases are easily generated.[27] That *it would be good for something or someone to be good* does not entail (or even reason to think, so it seems) *that it or she is good*.

---

[26] As I say, this is true on a standard understanding of possibility. And S4 and S5 versions of modal logic both capture this way of thinking. But it is not the only understanding available. On others, in which the accessibility relations that determine which worlds count as possible are not transitive the entailment does not go through. For discussions of possibility, then, the issue becomes whether one way of understanding possibility or another captures what we are claiming in saying something is possible in the context in which we are interested.

[27] See Rob van Someren Greve's "Wishful Thinking In Moral Theorizing: Comment on Enoch," *Utilitas*, vol. 23, no. 4 (2011), pp. 447-450, for arguments along these lines. There are epi-cycles we might go through here, say by wondering whether any alteration for the better in one person's character will be more than off-set by a moral loss somewhere else. But this would just be a distraction, since the problem with the inference doesn't turn on the particular facts but with the intelligibility of the case. Van Someren Greve considers the possibility that if we limit *p* to something that is itself a moral claim, say "that human happiness is good," a version of the inference from GGp to Gp might go through. As van Someren Greve points out, it is not at all clear that the restriction would facilitate the inference, and even if it did, it would leave us with a claim about a moral claim –

This version of a counter example doesn't really work, though, since 'good' is not being used consistently as an operator. Unfortunately, in English, it is a bit hard to get an example going that sounds grammatical. ~~Although I am hopeful that I can.~~ As a first step, really to prepare the ground for an example relying on *it is good that…* let me switch from talk of "it is good that" to talk of 'it is obligatory that….' Consider the suggestion that O(Op) → Op (where "O" stands for "It is obligatory that…"). Counterexamples seem easy to generate. For instance, suppose that it is obligatory that it is obligatory that Alex visits Barbara, say because (for some reason) it is obligatory that Alex promise Barbara that he will visit her (in a context where making that promise will make it true that it is obligatory that he keeps it). In such a case we cannot reasonably conclude that it is obligatory that Alex visit Barbara, since he may well not have made the promise and it might be that whatever reason there was to make the promise was not, independent of the promise, enough to establish that it was obligatory that Alex visit Barbara (nor was there any other reason sufficient to establish that). The same would be true if in place of the second obligation relying on a promise (one that turns out not to have been made), it relied on you having granted someone permission to do something, which you ought to have done, but did not, or on you preparing the circumstances in some way, which you ought to have done, but did not.

Nothing particular about promises is in play here. What matters is just that many obligations we have (or would have) depend on certain, perhaps morally required, conditions, which might not actually be in place. Many moral obligations exist only because of what we do, or fail to do, and unless we assume we always do as we ought, we will not be able, in any unrestricted way, to infer from *it ought to be that it ought to be that p*, that *it ought to be that p*. After all, it may be that we failed in satisfying the first ought and so undercut the presence of the second.[28]

Similarly, it seems, many things are good only because other good things have happened. It is good, for instance (let us suppose) to express sincere appreciation when someone does something nice for you (which, we may assume, it would be good for them to do). In such a case, it is or would be good that doing so is good (not least because it would reflect that someone has done something nice for you). Thus *it is good that it is good that p* (where p is you expressing sincere appreciation). That does not mean that it is actually good for you to express sincere appreciation; whether it is depends (let us assume) on someone having done something nice for you, and the person might not have, even if it would have been good had she.

---

that it is or would be good if human happiness were good, but not with the moral claim that we swapped in for *p*: that human happiness is good. See p. 450.

[28] Perhaps it is worth noting that in standard deontic logic "it is obligatory that…" is treated as formally equivalent to "it is necessary that…" relative to morally ideal worlds. The idea is that it is true that "it is obligatory that p" if and only if p is true in all morally ideal worlds. Deontic necessity, so conceived, does support an inference from OOp to Op, since everything that ought to be is in morally ideal worlds. But deontic necessity is not a model that fits well with what we are talking about in claiming something is obligatory (or permissible or good) in a less than ideal world like ours. David Lewis has a lovely discussion of issues in this area in "Rights to Rights," in *Theoria*, 69 (2003), pp. 160-165.

In whichever way these examples work out, there are two things to keep in mind. One is that if (what we might call) the self-establishment thesis is true of some moral operators – so that the goodness of something being good establishes that it is good, or that something's being such that it is obligatory that it is obligatory establishes that it is obligatory – we need an explanation of *why* these operators work in this way. The other is that any such explanation needs to take account of the fact that the form of argument appears to go through not just in cases where it would be good (or it ought to be) that certain things are good (or ought to be), but also in cases where it is good (or it ought to be) that certain things are bad (or forbidden).[29].

An alternative explanation, developed by Ryan Preston-Roedder, follows up on Nagel's suggestion that "one morality is more likely to be true than another, because the former makes for a better world than the latter — not instrumentally, but intrinsically." As Preston-Roedder articulates the "better world" argument it appeals to morality requiring that we have, as a "substantive moral aim" making the world better. Any plausible moral theory, he suggests, recognizes this as (at least one among) our important moral aims (though plausible moral theories differ quite significantly concerning what would make the world better). He then imagines two plausible moral theories that differ in the following respect: if the first is true the world is better place (thanks certain values are instantiated or respected) than if the other is true. (The truth of either he points out would be the realization of an important value: the articulation of a true theory of morality. So they are equal on that score.) But, he argues, the fact that the truth of the second would mean the world is less good than were the first true, is reason (albeit over-ridable reason) to believe that the first *is* true. This is because, as he thinks of it, the second theory would be self-defeating in this sense: it advances as an important substantive aim something that it, itself, fails to promote or achieve:

> it both assigns certain substantive aims to us and includes principles the mere truth of which would render those very aims worse achieved than the truth of other, rival principles.[30]

In making his case that this contrast among the theories constitutes an argument in favor of the truth of one of the theories, he explores two arguments: Slote's for moral autonomy and mine against moral dilemmas. Both are, Preston-Roedder thinks, plausible arguments and both turn, he thinks, on maintaining that their force (to the degree they have any) turns on an appeal to alternative theories being self-defeating.

On Preston-Roedder's reading, the two arguments work, if they do, by showing that

---

[29] There seems to be no easy way to define what is bad in terms of something else being good. This means it would be hard for the self-establishment thesis to play a role in explaining an argument that compared two theories of what is bad and argued in favor of one, against the other, on the grounds that it was a theory of a (morally) better morality. Things are easier if the theories are of what is forbidden and one is advanced on the grounds that it morally ought to be that certain things are forbidden, since the claim that things ought to be forbidden can plausibly be translated into claims that their negations ought to be.

[30] p. 634

any theory that denies their conclusion assign "certain aims to us" (to make the world better) but include "principles the truth of which would render these aims worse achieved than the truth of alternative principles" (principles that allow moral autonomy or rule out the possibility of dilemmas).[31]

Importantly, Preston-Roedder is not committed to the success of either argument. His interest is in explaining why the *form* of argument they share is credible. And his explanation is that they rightly take a certain kind of self-defeat (which they accuse other theories of suffering) as grounds for thinking a theory false.

As it happens, I am a fan of both arguments, Slote's appeal to the importance of a certain sort of autonomy as grounds for rejecting consequentialism and my appeal to the unfairness of demanding the impossible as grounds for rejecting moral dilemmas. But I find the appeal to *making the world better* as a substantive aim interestingly different than what I have in mind when I think of the plausibility of these arguments. Putting that aside, though, I am puzzled by the suggestion that a moral theory would count as self-defeating if it failed to promote the aim it specifies for us (and a true moral theory recognizes). For *self-defeat* to come in I am thinking the aim must also be properly specified as an aim of the theory, or perhaps an aim for us in accepting a theory. But neither strikes me as right. So, while I do think the moral credentials, so to speak, of the principles of the theory are at stake, I don't think the challenge is one posed by self-defeat.

Moreover, my sense is that what is important in the arguments is not a matter of *the world* being better or worse, instrumentally or otherwise, but of *morality* being so – specifically of morality being more (rather than less) justified, by its own lights. It may well be that by the standards of the (morally) better moral theory the world is in fact worse morally than the alternative theory would have it.

An alternative…

With that in mind, I would like to turn back to Enoch's suggestion that some moral terms express operators that establish their own application – that, for instance, if it would be good that it was morally good that something were morally good, that thing is morally good.

I think the truth lies closer to -- though only closer, not in -- a defense of the inference running the other direction, from something being good (or right or obligatory or justified) to it being good that it is (i.e., $Gp \rightarrow GGp$). The (more than slightly misleading) idea is that there must be something good about – some justification of – something *counting as good* (or bad or right or wrong or virtuous or vicious by a candidate standard) if it is actually good (or bad or…).[32] Thus, again a bit misleadingly: GGp is a necessary condition, not a sufficient condition, for Gp. The key dynamic revolves around the importance of not

---

[31] p. 637

[32] And something good about something counting as bad if it is actually bad, and something good about something counting obligatory, or forbidden, or permissible, if it is.

holding sticking to a standard of morality in a context where one has reason to think that there is a morally better standard.

If I am right, in each of the cases mentioned above, the force of the argument that one morality is morally better than another is that it gives us reason to suspect that the original theory (which we initially took to be the best theory of morality) is of a morality that is not, given the alternative(s), one that merits either our allegiance or compliance.[33] The arguments work by calling into question a credential that the actual standards of morality must have: justifiably making a claim on our allegiance.

One way to see the force of the demand that the standards of morality stand as themselves morally justified, is to consider a familiar dialectic concerning dispositional accounts of value.[34] According to such accounts, what it is for something to be valuable is for it to be such that it would garner approval (or some other positive response) from someone appropriately situated.

Such accounts face a familiar challenge: that to be plausible, whatever garners the requisite approval must merit the approval it garners.[35] According to many people, this challenge reveals that such accounts cannot in the end be adequate -- since they will need (these people suppose) a distinct standard of value to determine which things merit (and which do not) the approval the might garner.

This is too quick, though. Nothing precludes dispositional accounts redeploying the standard on offer (as set by the reactions of the impartial spectator) to determine whether what garners approval merits that approval; the test is simply whether their garnering approval would itself garner approval. If so, these views hold that the objects that would garner approval merit it – that it is good that they would receive it – and if not, they hold that the objects do not, after all, merit the approval they would garner.

This is not a trivial test. After all, people, properly situated, might well disapprove of their own pattern of approval. Nonetheless, it is also a test that might well be passed and if it is, those who embrace the account will be in a position to affirm that what garners the approval of those properly situation merits the approval garnered.

Thus, recognition of the requirement that, on such accounts of value, what garners approval must merit the approval does not mark a decisive objection to them. Yet it is a requirement that must be passed if the theory is to be plausible, and it is worth noting why. If a theory of value did not pass, it would mean that – by the theory's own lights – there was

---

[33] A bit more carefully. What the arguments purport to show that in certain respects the standards on offer by the putatively best theory are one's that on balance, given the alternatives, we do not have reason to comply with. We may well have reason to comply with many of the standards under most circumstances. To the extent this is so, the theory of a better morality will preserve those aspects of the original theory.

[34] Variations on this theme were advanced by Francis Hutcheson, David Hume, and Adam Smith, and more recently by Roderick Firth, Michael Smith, and David Lewis, among others. Include cites….

[35] John McDowell has made this point especially forcefully in "Values and Secondary Qualities,"…

nothing good about the standard it offered for distinguishing what was good from what not.[36]

Significantly, in accepting the requirement there is no assumption that passing the it is positive grounds for thinking a standard is the right one (so no commitment to thinking that GGp → Gp). Rather, there is the recognition that if a standard does not pass, it should be rejected (since, loosely and a bit misleadingly, Gp → GGp, and failing the test is a matter of discovering reason to think that, on the account in question, ~Gp).

The same line of argument, it is worth noticing, finds a home in discussions concerning theories of practical and theoretical reason.

Thus many argue that the standard maximizing theory of practical reason (according to which the rational choice is the one that maximizes expected utility) needs to be revised in light of its own focus on optimality as a measure of standards of reasoning, since maximizing expected utility predictably results in suboptimal outcomes in various game theoretic situations.[37]

Similarly, many people hold that an acceptable theory of induction needs to endorse itself as a measure of what to expect – on pain, otherwise, of judging its own measure of rational expectation irrational.[38]

In much the same way, others have argued that an acceptable rule for determining (partial) rational credence must by your own lights vindicate your actual credences.[39]

In each of these cases, the argument is that an acceptable normative theory in the relevant domain must not be one that fails to meet the justificatory standards it sets. Again, satisfying this requirement does not establish a normative theory as right, but failing to meet it rules a theory out as mistaken.

---

[36] Adam Smith criticizes Hutcheson for failing to recognize the importance of this requirement. See The Theory of Moral Sentiments, p. 323. Hume meanwhile makes a point of emphasizing that "not only virtue must be approv'd of, but also the sense of virtue: And not only that sense, but also the principles from whence it is derived" (in The Treatise of Human Nature, p. 619).

[37] See, for example, David Gauthier's defense of constrained maximization. Cite.

[38] See David Lewis on the requirement of immodesty in "Immodest Inductive Methods," *Philosophy of Science*, vol. 38, no. 1 (1971), pp. 54-63. His conclusion: "immodesty — in the class of otherwise adequate methods, under an appropriate accuracy-measure, on the total evidence — is a necessary condition of inductive adequacy" p. 63.

[39] See James Joyce's "A Nonpragmatic Vindication of Probabilism," *Philosophy of Science*, vol. 65, no. 4 (1998), pp. 575-603. He argues that only standards that satisfy the laws of probability theory will meet this requirement and takes this as a vindication of probabilism. J. Predd, R. Seiringer, E. Leib, D. Osherson, H. V. Poor, and S. Kulkarni, in "Probabilistic Coherence and Proper Scoring Rules" offer a proof that "No matter what proper scoring rule is in force, any incoherent forecast [that is, any forecast that is inconsistent with the probability calculus] can be replaced by a coherent one whose [epistemic] penalty is lower in every possible circumstance; there is no such replacement for a coherent forecast" p. 3. The quasi-dominance strategy in play here is analogous to the one I use in my argument against moral dilemmas.

<u>Tentative Conclusion</u>

The fact that the general style of argument reaches beyond morality recommends finding an explanation for it that does so as well. My hunch is that our normative concepts (whether moral or not) are concepts that make a claim on our allegiance, purporting to be such that their satisfaction *ipso facto* means – and here I purposely use a generic characterization that reaches beyond morality – someone or other has reason to act or react in certain ways to what satisfies them. Just who has the reasons, and what they have reason to do, varies dramatically according to the concepts in play. Thus our concept of duty, when satisfied by an action available to someone, might imply that the person has decisive reason to perform it, while our concept of value, when applied to some object, might imply that someone (just who might vary) has some reason to cherish, or promote, or protect it, and our concept of evidence, when applied to some consideration, might imply that people aware of the evidence have some reason to believe what it supports.[40]

This general feature of our normative concepts (moral and otherwise) crucially distinguishes them from non-normative concepts. Non-normative concepts may often be such that their satisfaction in fact provides reason for someone or other to act or react in some way. But such concepts do not purport to be such that their satisfaction *ipso facto* means that there are reasons. The reasons that the satisfaction of non-normative concepts provides, if indeed there are such reasons, depend on other (conceptually, though not necessarily metaphysically) distinct features of the situation.[41] Thus whether we have reason to stop at a particular red light depends on more than whether it is red, for instance on whether we have a desire to conform to the law, or a concern not to get a ticket, or… whereas our having (some) reason to perform a particular action is settled if it is valuable to do (or so our concept of value implies).[42]

Moreover, this feature of our normative concepts means that competence with them carries a distinctive burden: that we be sensitive to whether the criteria we rely on in applying such concepts are such that satisfying the criteria *ipso facto* establishes that someone or other has reason to act or react in a certain way. Evidence to the contrary – evidence that satisfying the criteria we are relying on does not *ipso facto* establish reasons – constitutes a direct challenge to the criteria. Such evidence gives us grounds for revising our criteria. That is what is going on, I think, when we discover a theory of a better morality, or a theory of better practical reasons, or a theory of a better principle of induction, or a theory of a better

---

[40] There is of course plenty of room for argument about which sorts of reasons, for which sorts of actions and reactions, of which people, distinguish one normative concept from another.

[41] It may well be that the very thing picked out by a non-normative concept – say the concept of PAIN – also always, and perhaps necessarily, satisfies a normative concept (e.g. BAD). Many think this is true of this example, though I do not.

[42] Needless to say, while (if I am right) normative concepts are such that their satisfaction establishes reason for someone to act or react, particular normative concepts may prove bankrupt, in ways that meant they are not actually ever satisfied. This is the view one might well take of, for instance, chastity in the eighteenth century. On this view, either chastity is a non-normative concept and the mere fact that some act is chaste gives no one reason to do act or react in any particular way, or else it is a normative concept the satisfaction of which requires that someone has such a reason, and no one does (so nothing actually satisfies the concept).

rule for determining what we should believe in situations of uncertainty. In each case, that the alternative is better in the relevant way provides grounds for thinking our original criteria (as captured by our current best theory of the domain) do not actually establish that we have the relevant reasons, at least when those standards conflict with those identified by the theory of a better set of criteria.

Another way to look at what I take to be the same phenomena, is to see that competence with our normative concepts, which purport to establish reasons, commits us to seeing arguments that there are relevantly better criteria for their application as grounds for thinking that the criteria we have been using need to be rejected or revised. , are not as a challenge to our understanding of the concepts in question.[43]

There are two ways we might think of the situation. On one, and this is how I have just put things, we are committed to seeing the standards we embrace (as criteria for our normative concepts) as themselves justified. On the other, and this is how I put things earlier, we are committed to holding that our normative concepts are such that their satisfaction *ipso facto* establishes reasons (for someone or other do act or react in some way to what satisfies them). In any case, I suspect these are actually deeply connected: we have relevant reasons to act or react in certain ways to the satisfaction of a concept only if the standard it sets is itself a justified standard for determining what reasons we have.[44]

A third way we might think of things, when focusing on morality, is this: our concept of morality is a concept of a system of rules, principles, and values than which none morally better can be conceived.[45] To find a set that is morally better than the set we have been thinking constituted morality is to find reason to think we have been wrong.[46]

---

[43] It is worth noting that arguments purporting to show that the fact that something is red give us no reason to do anything constitute no challenge at all to our understanding of the relevant criteria for redness – even if they might show that we shouldn't care about what is red.

[44] In thinking about whether certain standards are justified we face two familiar views. On one view: the standard is justified only if there is positive support for it. On the other view: the standard is (we might say permissively) justified as long as there is not, on balance, reason against it. The first invites a familiar infinite regress; the second avoids that, but traffics in permissive justification. I am a devotee of the latter, for a broad range of reasons. Right now, though, the point is that either conception of what it takes for a standard to be justified leaves us thinking that, when there is a relevantly better alternative, the standard is not justified. Alternatively, though, my point may be put this way: that our normative concepts are concepts that purport that their satisfaction ipso facto means that there is reason for someone or other to do (to act or react) in particular ways. The challenge posed by there being a theory of a better morality is that the standards of that morality seem to be the ones the satisfaction of which provide people with reason to act and react in various ways.

[45] Unlike a similar claim concerning our conception of God, about which one might worry that the morally best being we might conceive of might not exist, there is arguably, with morality, no slip between being conceived and being. On such an admittedly contentious view, morality's existence as a standard for us requires nothing more than being the set of standards than which none morally better can be conceived: there is no extra question of existence, as there is for God.

[46] Generalizing from morality to all normative domains, the idea would be that our concepts of normative domains are a concepts of systems of rules, principles, and values than which none relevantly better can be

conceived. To find a set that is relevantly better than the set we have been thinking constituted the domain is to find reason to think we have been wrong.