Philosophers' Imprint

VOLUME 15, NO. 4
JANUARY 2015

Non-Consequentialism Demystified

Howard Nye
University of Alberta
David Plunkett
Dartmouth College
John Ku

© 2015 Nye, Plunkett, & Ku
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.
<*www.philosophersimprint.org/015004/*>

1. Introduction[1]

Many moral theories hold that what an agent morally ought to do is determined in part by the goodness of the outcomes of those actions currently available to her. However, some moral theories go further and claim that what an agent morally ought to do is determined *solely* by the goodness of the outcomes of her actions. Following one standard philosophical convention, we will call this stronger thesis *consequentialism*.[2]

It seems quite clear that bringing about good outcomes is one thing that there is moral reason to do. But why think that promoting the good is the only thing that morality prescribes? One way of defending this thesis is to employ substantive normative arguments, which seek to show that the plausibility of moral reasons to do other

1. This paper has benefited from many conversations and from written feedback on many previous drafts. We would particularly like to thank Liz Anderson, David Braddon-Mitchell, Jonathan Dancy, Stephen Darwall, Tom Dougherty, Andy Egan, Mylan Engel, Robert L. Frazier, Allan Gibbard, Bob Goodin, Paul Hurley, Nadeem Hussain, Frank Jackson, Christine Korsgaard, Ira Lindsay, Dustin Locke, Tristram McPherson, Timothy Michael, Alastair Norcross, Doug Portmore, Peter Railton, Ryan Robinson, Tamar Schapiro, Jeff Sebo, Scott Shapiro, Michael Smith, Nic Southwood, David Velleman, Ralph Wedgwood, Caroline West, and two anonymous readers for *Philosophers' Imprint*. Earlier versions of this paper were presented at the University of Michigan, Ethics Discussion Group (June 2007), The First Annual Rocky Mountain Ethics Congress at the University of Colorado Boulder (August 2008), and The Twelfth Annual Oxford Graduate Conference in Philosophy at the University of Oxford (November 2008). Thanks to all those who offered us feedback during those presentations.

2. While this description may suggest that only theories that employ an impartial or agent-neutral kind of goodness can count as "consequentialist" in our sense, it will become clear how our arguments extend to relativized versions of consequentialism (of the kind defended by Sen 1983, Dreier 1993, Smith 2003, Louise 2004, and Portmore 2011). This use of 'consequentialism' does restrict its referent to "direct" or "act-consequentialist" theories, according to which the moral status of our acts is determined exclusively by the goodness of *their* outcomes. While these are the theories on which we will be focusing, we will also have something to say about "indirect consequentialist" theories, according to which the moral status of our acts is determined exclusively by the goodness of the outcomes of something else (like everyone's accepting or acting on a set of rules that ranks our acts in certain ways).

things does not withstand reflective scrutiny.[3] Many such arguments begin by observing that it is overwhelmingly plausible that there are moral reasons to do what is good for others and to avoid doing what is bad for them. These arguments concede that it is initially plausible that there are moral reasons to do other things, like keep our promises and respect the autonomy of others, even when this does not bring about the best overall consequences. Further, it is initially plausible that reasons not to harm others are stronger than reasons to benefit them, and that reasons to benefit our loved ones are stronger than reasons to benefit strangers. Such considerations do initially count against consequentialism. But, proponents of these consequentialist arguments contend, there are good reasons to doubt that the plausibility of these distinct reasons and weighting principles can withstand our getting clearer about what exactly they could amount to. If this is correct, our plausible moral reasons of beneficence and non-malfeasance are best seen as instances of moral reasons to bring about the overall best outcomes, which exhaust the content of morality. We will call this the "bottom-up" strategy of arguing in favor of consequentialism.

A second way of defending the consequentialist thesis, often pursued in tandem with the bottom-up strategy, is to maintain that consequentialism enjoys a kind of general theoretical advantage that stems, not from the failure of non-consequentialist considerations to withstand scrutiny, but from general reflections on the nature of morality and reasons to act. While it is controversial whether there is necessarily reason for everyone to be moral, it seems clear that morality is something *important*, in the sense that there is strong practical reason for most people to be moral most of the time.[4] It therefore seems desirable for a moral theory to be able to explain why there is reason for us to be moral by identifying the content of morality with something which general facts about the nature of practical reasons entail there is practical reason to do. There seems to be an inexorable connection between the thought that a state of affairs is good and the idea that there is reason for us to bring it about. While there may be other equally plausible thoughts about what there is practical reason to do, like the idea that there is reason to promote one's own interests or satisfy one's desires, none of these things could plausibly be identified with what there is *moral* reason to do. Thus, the consequentialist view that morality prescribes that we bring about good states of affairs not only seems substantively plausible but offers an explanation of why there is reason to be moral. We will call this the "top-down" strategy of arguing in favor of consequentialism.[5]

3. See for instance Sidgwick 1907 (esp. Bk. III, ch. xi, and Bk. IV, ch. ii–iii); Bennett 1966, 1981, 1995; Nielsen 1972; Kagan 1989; and Norcross 2003.

4. As we will use the phrases, "practical reasons to do *A*" (or just "reasons to do *A*" with no further modifier) are considerations that count in favor of doing *A* in the most general normative sense, while "moral reasons to do *A*" are considerations that contribute to *A*'s deontic status as morally permissible, right, or good to do. We can understand the question 'Why care about morality?' as asking "Why should we take moral reasons to be genuine practical reasons?" and the question 'Why be moral?' as asking either the foregoing or "Why, just because the moral reasons in favor of doing something morally require us to do it, should we think there is decisive practical reason to do it?"

5. A number of prominent advocates of consequentialism employ some version of the top-down strategy. For example, consider the case of Sidgwick (1907). Since Sidgwick was concerned not merely with what there is moral reason to do but with what there is practical reason to do (see Bk. I, ch. iii), we can read his argument for the "maxim of Benevolence" in Bk. III, ch. xiii, as holding that, just as it seems self-evident that there is practical reason to aim at one's own greatest good, "the good of one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other …. And as a rational being I am bound to aim at good generally … not merely at a particular part of it." So, Sidgwick suggests in his concluding chapter, if our moral reasons (or, in his terminology, our other-directed or "social" duties) exclusively favor promoting the "universal Good", practical reason will endorse moral reasons (even if it also threatens to contradict itself by giving an equally strong endorsement of conflicting considerations of self-interest). Similar (if more "instrumentalist") arguments are given by Smart (1956, 353), Singer (1979, 10–13 and 212–220), and Railton (1986, 189–190 and 202–204). The logically strongest version of the top-down strategy defends a teleological conception of practical reasons [TCR], according to which *all* reasons to act must be reasons to bring about outcomes that are good (from some perspective, even if not from an impartial perspective). For a sustained argument along these lines, see Portmore 2011 (although Portmore does not put his favored version of TCR or consequentialism in terms of reasons to bring about good outcomes, it will become clear from our account of evaluative

In this paper we argue that closer attention to *why* there are reasons to promote good outcomes shows the top-down strategy to be unsound. We argue that there are reasons to promote good states because these are states it is fitting to desire, desiring a state involves motivation to promote it, and it is a general conceptual truth that there are reasons to do what it is fitting to be motivated to do. But, we contend, there is every reason to think that some fitting motives aim simply at doing or omitting certain things, rather than promoting certain states of affairs. Indeed, we argue that an act's moral status consists in the fittingness of feelings of obligation to perform or avoid performing it, which motives are in the first instance directed at the act rather than the states it brings about. Thus, we argue, the same connection between fitting motives and reasons to act that explains why there are reasons to promote the good *directly* explains why there are reasons to be moral, *whether or not* being moral consists in promoting the good.

We believe that the unsoundness of the top-down strategy is important, not only because explicit uses of the strategy have been influential, but because many consequentialist arguments that look largely bottom-up gain some of their force by tacitly relying upon the top-down strategy. For instance, many of Shelly Kagan's (1989) arguments against non-consequentialist constraints on doing or intending harm seek to show that they cannot be clarified in ways that comport with the intuitions of those who appeal to them.[6] But Kagan goes on to argue that proponents of these constraints face a more important problem, namely that of *justifying* them or *explaining why* such factors as doing or intending harm should be so important. Kagan does not present the alleged need to independently justify non-consequentialist constraints as embodying any grandiose theory about the nature of practical reasons; he supports it by arguing that the mere fact that a moral theory fits our intuitions about what to do in particular cases cannot be a sufficient justification of it, else slaveholders would be justified in believing that skin color has intrinsic moral significance.[7]

But while it seems implausible in the abstract that skin color has intrinsic moral relevance, it actually seems — at least at the outset of inquiry — *plausible* in the abstract that inflicting harm is intrinsically worse than failing to provide aid and that setting out to harm someone is intrinsically worse than harming her as a foreseen consequence of one's conduct. Indeed, the intrinsic moral relevance of these factors seems — again, at the outset of inquiry — no less directly plausible than the idea that there is a general moral requirement to benefit others or promote the good. Of course, if Kagan can show that the direct plausibility of the intrinsic relevance of the doing / allowing and intending / foreseeing distinctions cannot withstand careful clarification, he will have made a compelling case against non-consequentialist constraints, and some of his arguments that "bring out the problems and unpalatable implications of adopting" them really do tend in this direction.[8] But why should we agree with the legitimacy of Kagan's demand that proponents of constraints provide an independent justification for why we should accept them *beyond* their direct plausibility *even if* this plausibility withstands the careful clarification of their content and how they apply to various cases? Nowhere in his book does Kagan provide a similar, direct-plausibility-independent justification of moral reasons to promote the good, or any reason to doubt that, if such a justification could be provided, plausibility-independent justifications of other moral factors could be just as easily provided.[9]

judgments why we think TCR and even relativized consequentialism can be roughly put in these terms).

6. See especially Kagan 1989, 87–111 and 128–165.

7. See especially Kagan 1989, 11–15, 111–127, and 165–182.

8. We have in mind especially his explorations of the implications of the doing / allowing distinction for decision making under risk (87–91); the possibility that the ways we draw the distinctions are gerrymandered functions of, and thus cannot justify, our intuitions about particular cases (101–106 and 138–144); and the particular difficulty of constructing a plausible story about the moral status of interrupting aid in progress (106–111).

9. Beyond, of course, simply providing some reasons to think that no factors other than promoting the good are genuinely relevant *because* their apparent relevance cannot survive critical scrutiny. But this does not support the asymmetric requirement that we must provide independent justifications of

What makes Kagan's inability to find successful plausibility-independent arguments in favor of non-consequentialist constraints look more like an argument for consequentialism than a skeptical challenge to our justification for holding *any* moral beliefs? It is, we think, the following common background assumption:

> **Mystification**: It is clear and obvious that there is moral reason to bring about the good, but mysterious how there could be moral reason to do anything else.

We think that quite a few arguments against non-consequentialist principles tacitly rely on the mystification assumption. Phillip Pettit's (1991) confident pronouncement that all moral theories must start by specifying what is good and then exhaust themselves by saying how we should relate to the good would seem entirely question-begging were it not that the good seemed to have a kind of unparalleled moral relevance. Similarly, the suggestion of Samuel Scheffler (1994, following Robert Nozick 1974) that there is an "apparent air of irrationality surrounding the claim that some acts are so objectionable that one ought not to perform them even if this means that more equally weighty acts of the very same kind or other comparably objectionable events ensue" (82) clearly seems to presuppose the incomprehensibility of moral reasons against any acts save those that fail to minimize the extent to which undesirable things happen in the world. (The kind of concern that Nozick and Scheffler are expressing is sometimes referred to as the "paradox of deontology".) To take one last example, Alastair Norcross (2003) offers several compelling arguments that the direct plausibility of an intrinsic moral difference between something like doing and allowing harm or negative rights against interference and positive rights to aid cannot withstand scrutiny. But he goes on to claim that, even conceding the unchallenged plausibility of the view that negative rights are weightier than positive rights, "we also need an explanation of … *why* [negative rights are] stronger than [positive rights]".[10] Yet Norcross does not seem to think that we need any similar explanation of *why* we morally ought to promote the good or minimize harm, evidently because he is mystified by moral reasons to do anything other than promote the good.

Why, then, *are* so many philosophers mystified by the idea of moral reasons to do anything other than promote the good? We think that it is a tacit sense that moral reasons to promote the good are supported by the top-down strategy's assertion that morality should provide us with genuine reasons to act, and that the promotion of good outcomes is the only sufficiently moral-looking thing that has a clear theoretical purchase on our reasons for action.

Our aim in this paper is to show that the top-down strategy fails, but that it is motivated by genuine connections among ethical evaluations, fitting motives, and reasons to act. Some authors have held that the very concept of a good state of affairs employed by consequentialists

non-consequentialist constraints *even if their plausibility withstood scrutiny* but no similar independent justifications of moral reasons to promote the good. Kagan is right that his stated aim of engaging with non-consequentialists whom he sees as already committed to moral reasons to promote the good relieves him of the need to provide an independent justification of these reasons (17–19). But he is wrong to think that the ad hominem nature of his arguments entitles him to use the inability of his opponents to provide plausibility-independent justifications of non-consequentialist constraints as an argument against them without providing plausibility-independent justifications of moral reasons to promote the good. This is because (a) the demand for such justifications may be entirely illegitimate, and (b) even if it is legitimate, we have no plausibility-independent reason to think that it will be more difficult to give these exalted justifications for non-consequentialist constraints than it will be to give them for moral reasons to promote the good, once we see what on earth these justifications are and how they can be given for anything.

10. Norcross 2003, 457–458. Like Kagan, Scheffler, and many others, Norcross speaks of an intrinsic moral distinction between negative and positive rights as having a great deal of "intuitive support", which is unfortunately ambiguous between (i) our having intuitions about cases that would be captured by the distinction, and (ii) its being plausible in the abstract that there is an intrinsic moral distinction of this kind. But context suggests that Norcross intends ii (perhaps in addition to i); for instance, he concedes the plausibility of a claim *the content of which* asserts the relative priority of the rights, namely, "My right not to be poisoned does seem stronger than my right, if any, to be given the food I need to survive" (457).

is either incoherent or devoid of the theoretically independent connection to practical reasons presupposed by the top-down strategy.[11] Against these authors we show how consequentialists' talk of good states is intelligible, and how there is indeed a deep theoretical connection between good states and reasons to act. But we argue that, unfortunately for the top-down strategy, once one understands *why* this connection holds, it becomes equally clear how there can be reasons — indeed moral reasons — to do things other than bring about good states.

In section 2 we vindicate the intelligibility and normative force of good states of affairs by analyzing them as states it is fitting to desire, and in section 3 we defend a conceptual connection between the fittingness of a motive and the existence of reasons to perform the acts it motivates. But while some of our motives are *state-directed*, or motives to bring about certain states of affairs, we contend that we have other motives that are *act-directed*, which are, in the first instance, motives simply to do certain things. In section 4 we show how, in the same way the fittingness of state-directed motives generates reasons to promote the good, the fittingness of act-directed motives generates reasons to do other things that may not promote the good. Moreover, we argue in section 5 that an act's moral status consists in the fittingness of a particular kind of act-directed motive, namely a feeling of obligation to perform or avoid performing it. This means that the same connection between fitting motives and reasons to act that explains why there are reasons to promote the good equally explains the connection between an act's moral status and reasons for or against performing it quite independently of whether the act promotes the good. This, we contend, demystifies how there could be moral reasons to do anything other than promote the good. We conclude in section 6 by examining how our argument may be extended to undermine the motivations for other theories that view moral considerations as subordinate to good outcomes and non-moral reasons more generally, like indirect consequentialism and contractualism.[12]

If our argument is successful, we believe that it removes an important source of support for consequentialism. Since, as we have indicated, many bottom-up arguments for consequentialism seem to rely tacitly on the top-down strategy, we think its elimination significantly weakens the bottom-up case for consequentialism as well. At the very least, we think that it forces consequentialists to demonstrate more directly why non-consequentialist principles are implausible without

11. See especially Taurek (1977, 304–306), Foot (1985, 202–209), and Thomson (1994, 12–14).

12. What is most central to our argument is the *existence* of the following connections between ethical categories, fitting attitudes, and reasons to act:

    (1) If something (*e.g.* state *S* or act *A*) falls under an ethical category (*e.g. goodness* or *moral wrongness*), then it is fitting to have a certain motivation towards it (*e.g.* to promote *S* or avoid performing *A*), and

    (2) If a motivation (*e.g.* to promote *S* or avoid performing *A*) is fitting, then there is reason to act as it would motivate us to act (*e.g.* to actually promote *S* or avoid performing *A*).

    In order to argue that these connections exist, we defend *particular explanations* as to why they exist, namely:

    (1*) Something's falling under an ethical category (*e.g. S*'s being good or *A*'s being wrong) *can be analyzed as* its being fitting to have particular motivationally laden attitudes towards it (like desires that *S* or feelings of obligation not to perform *A*), and

    (2*) *What it is for there to be* reason to perform an act (*e.g.* to promote *S* or avoid performing *A*) is for the act to constitute or achieve something that it is fitting to be motivated to do or achieve.

    We believe that much of our argument would remain sound if alternative explanations of (1) and (2) were correct — for instance, if the fittingness of attitudes were explained by the instantiation of ethical categories, or the fittingness of motives were explained by reasons to act. But we argue in favor of our particular explanations (1*) and (2*) because we believe that they are the best explanations of connections (1) and (2), and the explanations most conducive to our argument. In notes 21, 34, and 52 we explain how our argument could be made to work if one were to insist on explanations of connections (1) and (2) other than those we defend. We are grateful to two anonymous readers for *Philosophers' Imprint* for pointing out and encouraging us to discuss ways in which our main argument most centrally depends on (1) and (2), and can remain sound even if (1*) and (2*) are mistaken.

assuming at the outset that moral reasons to promote the good have a uniquely obvious rational sanction and demanding that, because non-consequentialist principles do not direct us to promote the good, we must give a special kind of justification for them.

2. Good Outcomes and Fitting Attitudes

To understand why the goodness of a state of affairs guarantees the existence of reasons to bring it about, we must begin by clarifying what it is to judge that a state of affairs is good. Such a judgment does not seem simply to describe the state. Rather, it also seems to involve a *normative* claim that recommends the state or speaks in its favor. In particular, it seems to be a claim to the effect that the state of affairs is desirable, or such that we should desire that it obtains. Put in terms of reasons, it looks like we might analyze the judgment that state of affairs *S* is good as one to the effect that there are sufficient reasons to desire that *S* obtains.[13]

While we think that this kind of analysis best captures the content and normative force of judgments about good states of affairs, it needs clarification and refinement concerning both the *kind of reasons* and the *kind of desires* it involves. Beginning with the first issue, there are clearly some kinds of reasons to desire a state that are of the *wrong kind* for constituting its goodness. Suppose, for instance, that an evil demon credibly threatens to harm your loved ones if you do not desire that there be an odd number of hairs on your head. The fact that the demon has made this threat seems to be a strong *pragmatic reason* to desire, or at least to get yourself to desire, that you have an odd number of hairs. However, this would not thereby make it good that you have an odd number of hairs.[14]

In contrast to pragmatic reasons, considerations like the fact that a state of affairs would involve children being happy do not simply count in favor of getting yourself to desire it — they make a desire for it fitting or appropriate. These *fittingness reasons* to desire a state that constitute its being fitting to desire it are the ones that constitute its goodness. An analysis of the concept GOOD STATES OF AFFAIRS into the concept REASONS FOR DESIRE should thus be in terms of fittingness, rather than pragmatic, reasons.[15] Of course, if the concept of a fittingness reason to desire a state just was that of a consideration that bears on whether the state is good, this sort of analysis would be circular. We think, however, that the distinction between judgments about fittingness as opposed to pragmatic reasons for attitudes can be made sense of without invoking ethical concepts like goodness.[16] While a full account of the distinction is beyond the scope of this paper, we think it can be usefully characterized by noting that it is characteristic of judgments about the existence of pragmatic reasons to have attitudes that they are *not* capable of directly guiding us into having those attitudes without first motivating us to do things to make it the case that we have them. To desire an odd number of hairs on your head in response to the reasons you take the demon's threat to provide, you must first do something like condition yourself, take drugs, or selectively attend to certain things with the aim of developing the desire. Judgments about the existence of fittingness reasons for an attitude, however, *do* seem capable of directly guiding us into having it without having to first motivate us to do anything to get ourselves to have it. For instance, judging that it is fitting or appropriate to desire knowledge for its own sake — as you might conclude upon contemplating

13. Where sufficient reasons to have attitude *A* are reasons that make one justified (in a sense we will clarify below), all-things-considered, in having *A*. For examples of this basic kind of analysis, see Ewing 1939, Gibbard 1990, Anderson 1993, and Scanlon 1998.

14. See Rabinowicz and Rønnow-Rasmussen 2004, following Crisp 2000. Some, like Gibbard (1990, 36), Parfit (2001, 27), Portmore (2011, 59), and Way (2012) insist that these pragmatic reasons *to get oneself* to desire something are not reasons *to desire* it at all. But whatever we want to call them, the important thing is to distinguish these kinds of reasons from the reasons to desire the state that (all agree) are connected to the goodness of the state..

15. In this paper, we follow one standard convention in the literature and use smallcaps to designate concepts.

16. For recent discussions of attempts to make sense of the distinction without invoking ethical categories, see, for instance, Danielsson and Olson 2007, Lang 2008, Way 2012, and Schroeder 2012.

Nozick's (1974, 42–45) experience machine — seems like it can directly cause you to desire knowledge intrinsically.

If we are to analyze judgments about a state's goodness as judgments that there are sufficient fittingness reasons to desire it, there remains a question as to *whose* reasons these are supposed to be. For instance, if both Jones's child and Smith's child are dying of kidney failure, and there is only one kidney available for transplant, it seems fitting for Jones to desire the state of its being transplanted into her child, but fitting for Smith to desire an alternative state of the kidney being transplanted into *her* child.

We think that the answer as to whose fittingness reasons one is talking about when one claims that a state is good is a contextualist one, the essentials of which have been suggested by Foot (1985), Hurka (1987), Lewis (1989), and Gibbard (1998). Clearly, when someone judges a state of affairs good, she ordinarily takes there to be fittingness reasons for *her* to desire it; the main question is who else's fittingness reasons she is making a commitment about.[17] Judgments that a state is good don't always seem to commit the judge to thinking that everyone has most fittingness reason to desire it. Surely Jones could truly say to her partner that their child's being moved ahead of Smith's child on the kidney waiting list was a good thing without being committed to thinking that it would be fitting for Smith to desire it. When Jones speaks to her partner about this being a good development, she seems to mean that *the fittingness reasons they share* support, on balance, a desire for it.[18] Since the fittingness reasons Jones shares with Smith include only basic reasons of beneficence to care about both children, equally unintensified by personal relationships to either, she could not truly claim that this development was good in talking to Smith.

Judgments that a state is good thus seem to be judgments that a desire for the state is supported by the fittingness reasons one shares with the group one means to be talking or thinking together with. In thinking about what she should want, it is possible for the judge's group to shrink to herself alone. It is also possible for the judge to address or think on behalf of the widest group of agents with whom she shares fittingness reasons to desire states. We think that this is the kind of context in which thinking about basic moral reasons of beneficence — to which all morality reduces, on some consequentialist views — takes place. To judge in this context that one state is better than another is to judge that the fittingness reasons one shares with all moral agents favor preferring it.[19]

This analysis of judgments about good states as judgments about our shared fittingness reasons for desire can defend talk about morally good states from the charges of incoherence or contentlessness levelled by Taurek, Foot, and Thomson. In asking someone to agree that it would be better if five others survived instead of her, we are not, as Taurek (1977, 305) suggested, claiming that she should, all things considered, prefer the survival of the five. We are merely asking her to agree that the fittingness reasons *she shares with the five* (and the rest of us) favor preferring their survival. Moreover, to claim in a universal

17. In certain contexts, such as those of giving advice, one might call states good that one does not take oneself to have fittingness reason to desire. For instance, in advising a rival in a competition, one might refer to states that tend towards her winning as "good", although one takes it to be unfitting for one to desire them oneself. One possibility is that, in giving such advice, the advisor takes on the false presupposition that there is reason for her to share the advisee's aims, and, given this, the advisor does in fact take there to be fittingness reasons for her to desire the states she calls "good". But another possibility is that, although the judge is almost always part of the audience she means to be addressing with talk about goodness, this fails to be true in special cases. Since the judge is clearly part of the audience in the contexts we will be considering, it will not matter for our purposes which account is correct.

18. To say that a response is supported on balance by a set of reasons is to say that there is no alternative response that those reasons favor more strongly.

19. This contextualist account thus provides a solution to "the Partiality Challenge" as to how we can analyze judgments about good states as judgments about fitting attitudes if Jones can judge it fitting for her to prefer that the kidney be transplanted into her child without thinking (in certain contexts) that this state is better (Ewing 1939, 19, and Blanshard 1961, 287–288). The solution is similar to Suikkanen's (2009), but it is more flexible in that it makes sense of non-relativized evaluative judgments in non-fully-impartial contexts, and it avoids the objections raised by Zimmerman (2011, 458–460) about circularity and indeterminacy.

context that state *S* is good is not simply to say, as Foot (1985, 202–204) and Thomson (1994, 12–14) suggest, that *S* has some particular property that one thinks we should promote. It is to make a substantive claim about what preferences among states are favored by the fittingness reasons we all share, which can be meaningfully debated.[20] Without such a fitting attitude analysis, it would be unclear what talk about morally good states could amount to, but with the analysis it becomes clear how this talk is of a piece with our other judgments about good states.[21]

20. We think many major debates in normative ethics concern what these reasons for preference support, including questions of what well-being consists in or which states we should prefer out of concern for a given individual (Darwall 2002); whether we should sometimes prefer lesser gains for some individuals to greater gains for other individuals in the name of equality (Temkin 2003), priority (Parfit 1997), or sufficiency (Crisp 2006); and whether we should prefer more individuals to exist simply because there will be more well-being in the world (Singer 1979).

    This said, there is a way in which we agree with some of Foot's (1985, 204–206) argument that *moral* reasons to promote morally good states must come from within, rather than outside of, morality, and that this makes trouble for what we are calling the top-down strategy of arguing for consequentialism. But, as we will argue in the next section, good states have a direct connection to practical reasons, which holds independently of any general connection between morality and practical reasons. We think this is exactly the sort of connection that the top-down strategy presupposes, and it does not, *pace* Foot, rely on any thoughts about what would be morally virtuous.

21. This is one reason why, although it is the existence of a connection between *S*'s goodness and there being fittingness reasons to desire *S* that is most central to our argument, we think our particular explanation of this connection is most germane to our argument. Suppose, for instance, that fittingness reasons to desire *S* could not be understood independently of *S*'s goodness (say, because what explained *R*'s status as a fittingness as opposed to a pragmatic reason to desire *S* was that *R* bears on *S*'s goodness). We could still give something like the foregoing contextualist account of judgments about *S*'s goodness, but it would have to be more like the following: (i) the judgment that *S* is good, when made in context *X*, has as its content the proposition that *S is good relative to X*, (ii) *S*'s being good relative to *X* entails that the fittingness reasons shared by the agents addressed in context *X* support on balance desiring *X*, and (iii) the judgment that *R* is a fittingness reason for agent *A* to desire *S* is true just in case *R* bears on *S*'s being good relative to X_A, where X_A is something like "*A*'s perspective.". While we think that our basic argument could be made using this alternative contextualist account, it would have to take as basic and unanalyzed the idea of "goodness relative to agent *A*'s perspective", which might be rather unclear (see *e.g.* Schroeder 2007 and Portmore 2011, 62). By taking the concept of a fittingness reason to be explanatorily prior to that of a good state, we avoid such problematic reliance on unexplained notions of perspective-relative goodness.

Having clarified the kind of reasons involved in our analysis, we now turn to the kind of attitude that judgments of a state's goodness take them to be reasons for. We characterized these roughly as reasons to desire that the state obtains, but this might seem too narrow. It can be awkward to speak of someone desiring states that currently obtain, that obtained in the past, or that are impossible for her to bring about, although she can of course judge (and judge truly) that current, past, and infeasible states are good. It is less awkward to say that we are *glad* that certain past and present states obtain and that we *hope* or *wish* that certain infeasible states will, and the thought that these responses are fitting seems to be what is involved in thinking them good. What being glad, hoping, and wishing that a state obtains have in common with desiring that it does is that they are "pro-attitudes" that involve an attraction to the state, which includes motivation to bring it about if one can, tendencies to find it appealing or pleasurable to think about, and tendencies to direct one's attention towards it and ways it might be realized.[22]

While some philosophers like to speak of every motivation and pro-attitude as directed at a state of affairs, we do not think that this is right. Consider the contrast involved in the following pairs of attitudes:

1. wanting to yell at someone in a fit of anger vs. wanting it to be true that one has yelled at him so he doesn't walk all over you,

2. an aversion to killing someone vs. an aversion to there being killings in the world, and

3. wanting to exercise now vs. wanting the world to be such that one exercises now.

22. For this sort of characterization of pro-attitudes, particularly in the context of desire, see Darwall 1983, 40–41; Scanlon 1998, 39; and Portmore 2011, 64–65.

Intuitively, the second attitude in each pair *does* take a state of affairs (one's having yelled, there being killings, one's exercising now) as its object. But the first attitude in each pair looks different; it seems to take a particular act (yelling, killing someone, exercising) as its object.

So says intuition, but why not simply treat the first attitude in each pair as a disguised motive to bring about a state of affairs, like "my yelling at him now" or "my exercising now"?[23] We think that state-directed and act-directed motives actually play rather distinct functional roles, which explain different aspects of cognition and behavior. A state-directed motive to bring it about that one does A now will explain both more and less than an act-directed motive to do A. For instance, a meek person averse to confrontation might have a strong state-directed motive to bring it about that he yells at a bully to get the bully to stop bullying him, but be unable to summon any act-directed motivation to yell at the bully. In such a case the state-directed motive will not by itself motivate the meek person to yell, but it will motivate him to do things that he hopes will cause him to yell, like ingest substances that he hopes will lower his inhibitions, direct his attention to considerations that might make him angry, or take anger-inducing pills if he has them ready to hand.[24] On the other hand, a customer who is angry with an online sales associate might have a strong act-directed motive to yell at the associate but think the anger is unwarranted and have no state-directed motive to bring it about that he yells at the associate. While the act-directed motive will motivate yelling and the various sub-actions that the customer takes to be ways of yelling (like moving his fingers in ways he takes to be the typing of angry messages the associate will see — which might not work if the Internet connection has gone down), it will not by itself motivate the customer to do things (like take inhibition-lowering or anger-enhancing pills) in order to cause himself to yell at the associate.

Thus, it seems that motivations to bring about states play the role of generating motivations to do things that are represented as bringing them about, while these motivations to do things (which may or may not be generated by motives to bring about states) play the role of generating motives to do more specific things that are represented as ways of doing the more general thing. While state-directed motives play the role of relating our intentional actions to states of the world, act-directed motives play the role of relating our more specific or local intentional actions to our more general or global intentional actions.[25] Consequently, we think that we can understand the distinction between act-directed and state-directed motives in terms of the following functional differences:

> M is an **act-directed motive** directed towards act A iff M plays the role of combining with representations that $<\phi$-ing is (or is not) part of a way of doing $A>$[26] to explain ϕ-ing,[27] while

23. See Portmore 2011, 56.

24. Of course, under happier circumstances, the meek person's motivation to bring it about that he yells would directly engender a motivation *to yell*, which would motivate yelling. What we illustrate is the causal work that the motive to bring it about that one yells can do even when it fails to give rise to a motivation to yell (which serves our aim of illustrating their different functional roles).

25. For a sustained discussion of how particular kinds of act-directed motives — namely plans or intentions — do this, see Bratman 1987.

26. By a "representation that $<\phi$-ing is (or is not) part of a way of doing $A>$", we mean something a bit broader than a representation that ϕ-ing actually constitutes doing A. For instance, an act-directed motive to yell at someone will motivate you not only to do things that you think will constitute yelling, but to do things like storming into the room next door where you think she is located in order to yell at her. While you presumably don't represent storming into the room as part of the act of yelling itself, the way in which the act-directed motive to yell gives rise to this motivation seems importantly different from the way in which a state-directed motive to bring it about that you yell can give rise to the motivation to take anger-inducing pills so as to cause yourself to yell. What we have in mind here is something like a generalization of the way an intention to A gives rise to more particular "intentions in action", or intentions to ϕ and thereby do A (as opposed to simply the way a desire for S gives rise to intentions to do A and thereby bring about S, even if S is a state involving one's performing certain intentional actions). For a review of literature on intention in action, see Wilson and Shpall 2012.

27. While we speak of act-directed motives to do A combining with representations that ϕ-ing is a way of doing A to explain ϕ-ing, one might worry whether

this is so in the case of basic actions, or actions one can perform directly and without performing any other intentional action. The analysis could be said to apply to basic actions because if A is a basic action, then an act-directed motive to do A combines with a representation that <doing A is a way of doing A> to cause one to do A. While this might appear artificial, this sort of application of the account might plausibly explain why individuals fail to perform basic actions that they want to perform if they don't know that they can perform them (*e.g.* if you want to move your arm, which has been paralyzed, and unbeknownst to you your basic ability to move your arm has just been restored, you may not move your arm, because you fail to represent to yourself that moving your arm is a way of moving your arm). Alternatively, one might say that the analysis we give in the text is correct so long as A is a non-basic action, but that if A is a basic action, the functional role of a motive to do A is simply to cause one to do A, all on its own and without any representations. Either way, we think that we will have no more difficulty accounting for motives to perform basic actions than those who reject act-directed motives and insist that all motives are motives to bring about some state S, which play the role of combining with representations that ϕ-ing will bring about S to explain ϕ-ing. Proponents of this teleological theory of motivation must treat motives to bring it about that you perform basic action A as either (i) combining with a representation that your performing A will bring it about that you perform A, which might seem artificial, or (ii) unlike other motives in that they are capable of causing one to do A all by themselves and without any additional representational states.

Another worry one might have about our claim that act-directed motives to do A combine with representations that ϕ-ing is a way of doing A to explain ϕ-ing is that such representations might seem too cognitively sophisticated to be necessary for an individual to have act-directed motives (and state-directed motives — since our analysis explains these in terms of their tendencies to give rise to act-directed motives). Young human children and many non-human animals are surely capable of motivations to do things, but they might seem to lack the concept of one action's constituting a way of performing another. We believe, however, that we face no more difficulty here than proponents of teleological theories of motivation, since it seems just as problematic to attribute the concept of an act's causing or bringing about a state of affairs to many motive-capable children and non-human animals. To explain how we can attribute such representations of constitution and causation to these individuals, we can argue that (i) the representations are implicit or *de re*, (ii) the required concepts of constitution and causation need not be as cognitively sophisticated as ours, or (iii) we overestimate the differences between our conceptualization of constitution and causation and theirs (perhaps because we confuse our thoughts about constitution and causation with meta-cognitive reflection or access to our thoughts about constitution and causation). For example, Dretske (1988) argues that discriminative instrumental conditioning involves implicit representations of the relevant kind of causation (109–122), observational learning involves explicit representations of such causation (137–150), and differences in the thinking about causation

> M is a **state-directed motive** directed towards state S iff M plays the role of combining with representations that <A-ing will bring about (or prevent) S> to explain act-directed motives to do (or avoid doing) A.

We will call pro-attitudes that are in the first instance attractions to states, which include state-directed motives to bring them about if one can, *state-directed pro-attitudes*. We will call pro-attitudes that are in the first instance attractions to performing particular acts, which include act-directed motives to perform them, *act-directed pro-attitudes*.

Having clarified this much, we can state our proposed analysis of the concept GOOD STATE OF AFFAIRS:

> **Fitting Attitude Analysis of Good States**: To judge that state of affairs S is good is to judge that the fittingness reasons one shares with a contextually specified set of agents, G_x, support on balance a state-directed pro-attitude towards S.[28]

of both cognitively less sophisticated and cognitively more sophisticated individuals may depend holistically on the networks of beliefs they happen to have and be largely a matter of degree (150–155). We believe that similar arguments can be made to support attributions of representations of the relevant kind of constitution to all motive-capable human children and non-human animals; we suspect that representations of the sort of constitution we have in mind are implicit not only in the general psychology of learning, but in particular in the functional roles that hierarchical action schemas are invoked to explain (see for instance Cooper and Shallice 2000 and Grafton and Hamilton 2007).

We are grateful to an anonymous reader for *Philosophers' Imprint* for raising these issues and pointing out other problems with our account of the distinction between act-directed and state-directed motives in an earlier draft.

28. This analysis leaves open a second way in which the truth conditions of judgments about good states depend on context, namely upon which alternatives to the state are relevant. Suppose that one of three mutually exclusive things might happen to the Jones's child: (n) she receives no kidney transplant and dies soon, (t) she receives a transplant but must take anti-rejection drugs and may have a shorter-than-average life, or (b) a medical breakthrough occurs which restores her own kidneys' functioning, ensuring a fully normal life. In a context where (n) was the relevant alternative to (t), the Joneses could truly judge that (t) is good, but in a context where (b) was the relevant alternative

We believe that this analysis has a number of important strengths. The first is its ability to explain the diversity of conceptually coherent (if often badly mistaken) views that people can have about what states are good, ranging from the judgment that the preservation of natural beauty is good to the judgment that a race's staying "pure" is good. Our analysis explains what is common to all such judgments, namely that those who make them think we should have pro-attitudes towards the states in question. A second strength of our analysis is its ability to capture the normative and motivational force of judgments about which states are good. To call state *S* good is to recommend *S* or speak in *S*'s favor, which we think is best captured by the claim that it is fitting for us to have pro-attitudes towards *S*. The analysis also explains the ability of judgments that states are good to motivate us to pursue them as a special case of the ability of judgments that attitudes are fitting to directly guide us into having them.[29]

To appreciate how central these features are to judgments about good outcomes, suppose that someone used the word 'good' to label all those states we call good, but took this to have no significance for what it was appropriate to desire and consequently had no propensity to desire the states in question. We suspect that the person would not really mean good when she said 'good'. On the other hand, if someone were to label as "good" precisely those states we think bad, we think she would be perfectly intelligible as thinking them good so long as she thought it was fitting to desire these states. That said, this fairly strong claim — that judgments that it is fitting to have pro-attitudes towards a state exhaust the content of judgments that it is good — could actually be weakened for our purposes here.[30] All we will rely on below is the claim that judgments about the relevant fittingness reasons are entailed by or *part* of the content of judgments that a state is good.

3. Fitting Attitudes and Reasons to Act

We have thus argued that to judge a state good is to judge that there are fittingness reasons that on balance favor our having a pro-attitude towards it. Since a judgment's truth entails the truth of its analysans, this means that it is a conceptual truth that a state of affairs *is* good iff *there are* fittingness reasons that on balance favor our having a pro-attitude towards it.[31] We will now argue that this analysis, together with general facts about the relationship between fitting attitudes and reasons to act, explains why a state's goodness entails the existence of reasons to bring it about.

The basic idea here is that what there is reason for us to do is determined by what aims there is reason for us to have (together, of course, with descriptive facts about what will achieve these aims), and the

to (t), they could truly judge that (t) is bad. Our analysis is consistent with this, so long as which states it is fitting for each agent to have *non-comparative* pro-attitudes (like desire, gladness, and hope) towards depends upon which alternative states are relevant in the context of judgment. What is presumably foundational are context-independent facts about which *comparative* pro-attitudes or preferences among states it is fitting for each agent to have (*e.g.* the Joneses should prefer (b) to (t) and (t) to (n)), and it is fitting for an agent to have a non-comparative pro-attitude towards a state just in case it is fitting for her to prefer it to the relevant alternatives (see Gibbard 1998). We are grateful to Doug Portmore for suggesting this example and encouraging us to discuss this issue.

29. Other plausible examples of this general propensity of judgments that attitudes are fitting include the abilities of judgments that beliefs are warranted by our evidence, that it is appropriate to be angry at someone, and that we should (in a non-pragmatic sense) intend to do certain things to directly generate those beliefs, feelings of anger, and intentions (see Gibbard 1990, 36–76; and Scanlon 1998, 18–22).

30. One might, for instance, think that judgments about what's good involve certain substantive platitudes or normative presuppositions about the sorts of things that can count. See Foot 1959, 85, and discussion by Gibbard 2003, 142–178.

31. Compare: if judging someone to be a bachelor amounts to judging him to be a male who isn't in a romantic relationship but in a position to enter one, then it's a conceptual truth that someone is a bachelor iff he's a male who isn't in a romantic relationship but in a position to enter one. Because analyses of one kind of judgment into another in this way support analytic relationships between the facts the judgments represent, we will slide rather freely between talking about what it is to make a certain kind of judgment ("to judge a state good is to judge it fitting to have a pro-attitude towards it") and talking about the analytic relationships between the facts they represent ("it's a conceptual truth that a state is good iff it's fitting to have a pro-attitude towards it").

question of what aims there is reason to have is identical to that of what it is fitting to be motivated to do. Since pro-attitudes like desiring or hoping that state *S* obtains involve motivation to bring *S* about (if one can), the fittingness of these attitudes entails the fittingness of this motivation, which entails the existence of reasons to bring *S* about (if one can).

The first part of this connection between fitting attitudes and reasons to act can be stated as a

> **Warrant Composition Principle** [WCP]: Let *P* be a psychic state that involves psychic state *P′* as an essential component. If *R* is a fittingness reason to be in *P*, then *R* is a fittingness reason to be in *P′*.

WCP simply states that if there is reason to be in a psychic state, then necessarily there is reason to be in all that the state essentially involves. For instance, if one acknowledges our claim that part of what it is to desire or hope that *S* obtains is to be motivated to bring *S* about if one can, it would seem incoherent to hold that a consideration (like *S*'s involving happy children) counts in favor of hoping or desiring that *S* obtains but does not count in favor of being motivated to bring *S* about if one can. Since having the motivation is simply part of what it is to have the pro-attitude, a consideration cannot make the pro-attitude fitting without making the motivation fitting as well.

The second part of this connection between fitting attitudes and reasons to act is the relationship between what it is fitting to be motivated to do and what there is reason to do, which we state as a

> **Motivations-Actions Principle** [MAP]: Let ϕ-ing be an action. If *R* is a fittingness reason to be motivated to ϕ, then *R* is a reason to actually ϕ.

Just as the consideration that *S* would make children happy can't make it fitting to desire *S* without making it fitting to be moved to bring *S* about when one can, so too it seems this consideration can't make it fitting to be moved to bring *S* about when one can without actually counting in favor of bringing *S* about when one can.[32]

It is intuitive that what there is reason to do is determined by what aims there is reason to have. We think that the best theoretical explanation of MAP is that, because practical reasoning governs our actions by means of governing our motives, the process of determining what aims to have — and thus what to do — is essentially a process of determining what intrinsic motives to have. As Michael Bratman (1987, 54) has suggested, because our practical reasoning controls our actions by controlling our intentions to perform them, reasons to perform an action just are reasons to intend to perform it (for a related idea, see Scanlon 1998, 20–21). But it must be clarified that reasons to do *A* are identical to *fittingness* reasons to intend to do *A*. As Kavka's (1983) toxin puzzle illustrates, merely pragmatic reasons to intend to do something (like the reason to intend — or get oneself to intend — to drink a toxin constituted by the fact that a reliable mind-reader will pay you if you intend this) need not be reasons to actually do it.

Moreover, in light of the role intentions play in realizing the objects of our desires and other pro-attitudes, there are similar reasons to think that, because reason ultimately governs our intentions by governing these attitudes, the fittingness of intentions is itself determined by the fittingness of these other motives. As Bratman argues, the role of intentions is not to supply an utterly new source of motivation that conflicts with the motives involved in our pro- (and con-) attitudes (like desires for states of affairs), but to help cognitively limited agents like us realize the objects of these motives over time. This role of intentions entails that their normative assessment must be tied closely to that of the pro-attitudes they serve.

Although Bratman often speaks as though practical reasoning must simply take our intrinsic pro-attitudes as given, it seems clear that

32. WCP and MAP closely resemble John Skorupski's principles FDF and FDD, the conjunction of which he referred to as the "Feeling / Disposition Principle" (1999, 38, 63, 131, and 174 n24) and more recently as the "Bridge Principle" (2010, 265–267).

we can assess them as reasonable or unreasonable by determining through philosophical reasoning whether they are fitting or unfitting.[33] Moreover, as we have seen, it is characteristic of these fittingness assessments that they directly guide our attitudes. For instance, one might start out with a much stronger intrinsic desire for one's co-nationals to be benefitted but, as a result of reflection on just what co-nationality amounts to, come to regard this preference as arbitrary. This judgment that one's preference is unfitting tends directly to change it, and to alter one's intentions from serving the old aim of benefitting one's co-nationals more to serving the new aim of benefitting aliens just as much.

Thus, because reason governs motives other than intentions through determinations of their fittingness, and intentions are simply a means of achieving the objects of these motives, fittingness reasons for intention are identical to fittingness reasons for these other motives. The role of intentions is primarily to enable us to settle in advance what future courses of action will best achieve the ends that it is fitting to be most motivated to achieve, and to pick from among the many courses of action that often have equally good prospects of doing this. Together with the above observation that reason governs our actions through determinations of the fittingness of the intentions that lead us to perform them, this entails that, because reason ultimately governs our intentions and actions by determining the fittingness of the motives they seek to serve, fittingness reasons to be motivated to do something are identical to fittingness reasons to intend to do it and practical reasons to do it.[34]

It is important to clarify that neither MAP nor this explanation of it makes what there is reason for us to do dependent upon what attitudes or motives we actually happen to have. The idea is that what there is reason for an agent to do is what would serve the objects of fitting intrinsic motives, by which we mean the intrinsic motives that it *would be fitting* for her to have, *whether she has them or not*. On this view, if a consideration (like a policy's benefitting the worse-off) is a fittingness reason for an agent to be motivated to do something (like vote for the policy if she can), then it is a reason for the agent to do this regardless of whether she ever has or comes to have any actual motivation to do it.[35]

33. While Bratman often speaks as though our intrinsic desires or pro-attitudes are *themselves* normative reasons for intention and action, he makes it clear that he actually wishes to remain neutral between this view and the view that our intrinsic desires can be assessed as reasonable or unreasonable (1987, 22).

34. As with the connection between a state's goodness and the fittingness of pro-attitudes towards it, what is most central to our argument is the *existence* of the connection asserted by WCP and MAP between the fittingness of motivational attitudes and reasons to act, rather than our particular explanation of why these principles are correct. Suppose we were wrong about the fittingness of motivational states being explanatorily prior to reasons to act. One might insist, for instance, that we cannot distinguish between fittingness and non-fittingness reasons for motivations and intentions without appealing to facts about which considerations are reasons to perform the acts that these motives and intentions would incline us to perform. While we think that our basic argument could be made using this alternative explanation of MAP, it might not fit as well with our claims about the relationship between fitting motives and ethical categories, or our claims to be giving a deeper explanation of the connection between ethical categories and reasons to act. One could not hold that what there is reason to do by itself explains the fittingness of all motivational attitudes, since both fitting and unfitting motivational attitudes (*e.g.* wanting *S* out of fitting care for someone and wanting *S* out of unfitting aesthetic appreciation for *S*) can incline us to perform the same acts. One might hold that something's instantiating an ethical category (like *S*'s being good) jointly explains the fittingness of pro- or con-attitudes towards it (like desires for *S*) and reasons to act in certain ways (like to bring about *S*). We think that we could still use this account to argue that the connection between good states and reasons to act should be seen as a special case of the connection between ethical categories, fitting motives, and reasons to act. But it would weaken the sense in which we would be explaining — as opposed to appealing to — the relationship between ethical categories and reasons to act. In part because of this, there would seem to be more room on this sort of "joint explanation" picture to hold that a state's goodness explains both the fittingness of desires for it and reasons to bring it about, but to deny that the instantiation of other ethical categories explains reasons to act in addition to the fittingness of certain motivational states.

35. The point of the appeals to how actions and intentions are governed by reason is simply to establish that, because (in the absence of something going wrong — as when we do the right thing for the wrong reasons) we can only respond to *genuine* normative reasons to act and intend by our motives first responding to these considerations, fittingness reasons to be motivated to do something are fittingness reasons to intend to do it and practical reasons to

Having thus argued in favor of our fitting attitude analysis of good states, WCP, and MAP, we can use them to give the following explanation of why, if a state of affairs is good, there is reason to bring it about:

1. If state of affairs *S* is good, then there are fittingness reasons for us to have a pro-attitude towards *S* [from the fitting attitude analysis of good states].

2. If there are fittingness reasons for us to have a pro-attitude towards *S*, then there are fittingness reasons for us to be motivated to bring *S* about if we can[36] [from WCP and the fact that pro-attitudes towards *S* essentially involve motivation to bring *S* about if one can].

3. If there are fittingness reasons for us to be motivated to bring *S* about if we can, then there are reasons to bring *S* about if we can [from MAP].

∴ If *S* is good, then there are reasons to bring *S* about if we can.

do it. The underlying idea is that if *R* is a genuine normative reason for us to respond in way *W*, and (absent something going wrong) we can only have *W* in response to *R* by having (and because we have) *W** in response to *R*, then *R* is a reason to *W**, and its status as such explains its status as a reason to *W*.

36. At least so long as it would still be fitting to have a pro-attitude towards *S* if you actually brought *S* about. Perhaps it is conceptually coherent to think that there are states that would not befit pro-attitudes if you did anything to bring them about. One might think it fitting to desire the existence of a natural environment, but that it would be unfitting to desire its existence if it were damaged and then restored, on account of its losing its naturalness. Even here, there would be cases where you could bring about the environment's existence without destroying its status as fittingly desired — *e.g.* by preventing others from destroying it or just refraining from destroying it yourself. Alternatively, one might think it fitting to desire some state, but that it would be wrong to try to bring it about, and consequently that it would be unfitting to have a pro-attitude towards it (qua product of wrongdoing) once it was brought about. Even here, we are not sure whether we should say that there would be *no* reason to bring about the state, or whether there would still be some reason to do so, which is simply outweighed by the features that would make it completely unfitting to have pro-attitudes towards it once it existed.

4. Rejecting the Teleological Conception of Practical Reasons

Recall that the top-down strategy of arguing for consequentialism contends that morality should give us reasons to act, that there clearly are reasons to bring about good states of affairs, but that it is unclear how there could be reasons to do anything else that looks moral. We have so far offered an explanation of the clear reasons to bring about the good that motivate the top-down strategy as a special instance of a general connection between fitting attitudes and reasons to act as they would motivate us. But in this section and the next we will undermine the top-down strategy by showing how this same general connection between fitting attitudes and reasons to act makes it equally clear how there could be reasons — and indeed moral reasons — to do things other than promote the good.

The logically strongest version of the top-down strategy holds that moral reasons to act must be reasons to bring about good outcomes because *all* reasons to act are reasons to bring about good outcomes. This view that all practical reasons must be reasons to bring about good outcomes is known as *the teleological conception of practical reasons* [TCR]. Of course, it certainly seems that there are reasons for each of us to do things like avoid our own pain or help our own children that are not proportionate to the badness of our pain or our children's distress *assessed from an impartial perspective*. That is, although there might seem to be stronger reasons to avoid our pain and help our children than to alleviate the pain and help the children of others, we could not truly say to those others that our pain or our children's distress is worse than theirs. But, as we have seen, in contexts (like talking to our partners) where our interlocutors share our fittingness reasons to have a stronger aversion to our pain or our children's distress than to that of others, it seems that we can truly claim that developments that mitigate the former are better than those that mitigate the latter.

TCR can thus accommodate practical reasons to do things other than promote the impartial good by holding that the practical reasons there are for an agent to act must be reasons to promote outcomes

that *she* can truly judge good. As we have seen in the last two sections, what determines both what states an agent can judge good and her reasons to promote such states are the fittingness reasons there are for her to have pro-attitudes towards them. So the clearest way to put TCR is as the thesis that all reasons for an agent to act are reasons for her to bring about states of affairs that there are fittingness reasons for her to have pro-attitudes towards.

Perhaps the most natural way of motivating TCR is to argue, as we have in support of MAP, that reasons to act are reasons to bring about the objects of fitting motives, and then to simply assume that all motives (and thus all motives that can be fitting) are state-directed. Thus Douglas Portmore claims, "If our actions are the means by which we affect the way the world goes, and if our intentional actions necessarily aim at making the world go a certain way, then it is only natural to suppose that what we have most reason to do is determined by which way we have most reason to want the world to go" (2011, 56). In claiming that "our intentional actions necessarily aim at making the world go a certain way", Portmore seems to be assuming that all motives on which we act aim at bringing about certain states of affairs.

We have argued that Portmore is right that intentional actions aim at achieving the objects of our motives, and that what we have most reason to do is determined by what we have most reason to be motivated to do. But we have also argued that Portmore's apparent assumption that all motives aim at states of affairs is mistaken. We suppose one *could*, as Portmore says, "intend to run merely for the sake of bringing it about that one runs" (2011, 56). But that would be bizarre. In cases where you run just because you feel like running, you have an act-directed intrinsic motive *to run*, and you form an act-directed intention to do this. Ordinarily, you do not seek to make it the case that the world contains instances of your running (or instances of your running *now*) unless you have ulterior reasons to ensure this (like your being paid for how much running time you log) and you find yourself without any act-directed motivation to run.

We think, moreover, that many of the attitudes we commonly have involve intrinsic act-directed motives. Consider emotions like anger, guilt, shame, admiration, contempt, grief, and aesthetic appreciation. Plausibly, these emotions essentially involve, respectively, intrinsic motivation to punish or hold the object of our anger accountable, intrinsic motivation to make amends for the object of our guilt, intrinsic motivation to withdraw and remove the object of our shame, intrinsic motivation to emulate the object of our admiration, intrinsic motivation to avoid and behave unlike the object of our contempt, intrinsic motivation to honor the object of our grief, and intrinsic motivation to engage with (*e.g.* to create, perform, watch, or listen to) the object of our appreciation.[37]

These attitudes that involve intrinsic act-directed motivations often seem fitting or appropriate, and there is no reason in the abstract to be any more skeptical that they *are* sometimes fitting than there is to be skeptical that state-directed pro-attitudes are sometimes fitting.[38] When these attitudes are fitting, the same two principles that explain why there are reasons to bring about good states of affairs will entail that there are intrinsic reasons to act as they motivate us, quite independently of whether our doing so will bring about good states. Since these attitudes essentially involve intrinsic act-directed motives, WCP entails that any fittingness reason to have them is equally a fittingness reason to have the act-directed motives they involve. Moreover, MAP implies that these fittingness reasons to have the intrinsic act-directed motives are actually intrinsic reasons to perform (or avoid performing) the acts they are directed towards (or against).

37. See for instance Gibbard 1990, Anderson 1993, Velleman 2002, and Darwall 2002, 2006.

38. It would, moreover, be self-defeating for teleologists to reject the idea of fittingness reasons and acknowledge only pragmatic reasons for attitudes. This is because their own account of the relationship between reasons for state-directed motives and reasons to act must be in terms of fittingness rather than pragmatic reasons for motives (else the pragmatic reason to intrinsically desire that your loved ones be harmed constituted by a demon's threat to harm them unless you form this intrinsic desire would count — absurdly — as an intrinsic reason to actually harm them).

For instance, suppose (as seems plausible) that the fact that your conduct has harmed someone makes for the fittingness of feeling (*i.e.* is a fittingness reason to feel) guilt for what you have done. Since guilt essentially involves intrinsic motivation to make amends, WCP entails that this fact is equally a fittingness reason to be intrinsically motivated to make amends, and MAP entails that this fittingness reason to be intrinsically motivated to make amends is actually an intrinsic reason to make amends for what you have done. Similarly, suppose the fact that a painting would look a certain way or a piece of music would sound a certain way makes for the fittingness of aesthetically appreciating the potential painting or music. Since aesthetic appreciation essentially involves intrinsic motivation to engage with its object (for instance by composing or performing it), WCP entails that this fact is equally a fittingness reason to be intrinsically motivated to engage with the painting or music, and given this, MAP entails that it is an intrinsic reason to actually engage with the painting or music (*e.g.* to actually compose or perform it). These *act-directed reasons* to act as fitting intrinsic act-directed motives dictate need no more sanction from fitting state-directed motives than *state-directed reasons* to act as fitting intrinsic state-directed motives (like that to promote someone's happiness) dictate need sanction from fitting act-directed motives.

Of course there are times when we acknowledge that an act-directed motive would be fitting, but find ourselves without it. We might feel no guilt for having harmed our rival or feel no grief at his passing, yet recognize that it would be fitting to feel some such guilt or grief. In *these* cases it is natural, and it seems fitting, for us to desire the state of our having these act-directed motives or at least that of our responding to the act-directed reasons constituted by the factors (*e.g.* that we have harmed our rival or that he has passed away) that make for their fittingness. Moreover, in cases in which we have and act from what we regard as fitting act-directed motives, it is natural and evidently fitting to be glad that we have and act from them. This is a special case of the general fact that it is natural and evidently fitting to have pro-attitudes towards our feeling and acting as it is fitting and reasonable to feel and act. But the existence and fittingness of pro-attitudes towards states of affairs in which we act as fitting act-directed motives would incline us (*e.g.* in which we make amends to our rival or attend his funeral) *are explained by, and thus cannot supplant* the fittingness of these act-directed motives and the basic act-directed reasons they generate.[39]

Portmore has argued that TCR should be accepted because it provides a maximally plausible and unified account of the relationship between which outcomes we have most reason to prefer and which acts we have most reason to perform. There clearly are *some* cases in which there is more reason for us to perform act a_i than act a_j because there is more fittingness reason for us to prefer the outcome of a_i (call this o_i) to the outcome of a_j (call this o_j). For instance, our stronger reasons to invest in a mutual fund that will deliver a higher return than its alternatives clearly stem from our stronger reasons to prefer the outcome of investing in the first to that of investing in the second. There are other

39. In response to Anderson's "expressive theory" of practical reasons that one should act only in ways that adequately express one's rational attitudes (which bears close similarities to our explanation of the relationship between fitting motives and reasons to act), Portmore (2011, 80) correctly observes that the teleologist can point to desires for states (like that of one's current actions adequately expressing one's rational attitudes) that would motivate the same acts as Anderson's theory. Moreover, it is plausible that these desires are fitting. But Portmore seems quite wrong to suggest that the fittingness of desires for these states *explains* our reasons to act as our fitting act-directed motives dictate. Our reasons to act as fitting act-directed motives dictate are no more in need of the fittingness of a desire that we act as our fitting motives direct than our reasons to act as fitting *state-directed* motives (*e.g.* to promote someone's happiness) are in need of the fittingness of this general desire. In both cases the connection between the fittingness of the motive and the reasons to act out of it are explained by WCP and MAP directly, and without the need of the fittingness of some general desire to act as there is reason to act (*i.e.* to act as fitting motives direct). The fittingness of a desire to promote someone's happiness [DH] and the state-directed reasons it generates to promote her happiness (together with the fact that it is fitting to desire that we feel and do what is reasonable) explain and are not explained by the fittingness of desiring the state of affairs in which one acts on DH. In the exact same way, the fittingness of an act-directed motivation to make amends [MA] and the act-directed reasons it generates to make amends explain and are not explained by the fittingness of desiring the state of affairs in which one acts on MA.

cases where our reasons to perform a_i over a_j do not seem to stem from our reasons to prefer o_i to o_j. For instance, when breaking a promise to do A will bring about slightly greater benefits than doing A, the mere fact that we have promised to do A can seem to be a stronger intrinsic reason in favor of doing A than the additional benefits are reasons to omit A. Here our apparently stronger reasons to keep the promise do not seem to be generated by reasons to prefer the state in which we keep it; rather our reasons to prefer this state seem to be generated by the fact that it involves our doing what there is greater reason for us to do. But, Portmore argues, proponents of TCR can re-describe our apparent intrinsic reasons to keep promises as reasons to bring about the state of affairs in which we keep our promises (or our current promises), and hold that these reasons do stem from our reasons to prefer the state of our keeping our promises. Although this may conflict with initial appearances, Portmore argues that we should accept it because it provides a more unified account of the relationship between our reasons for preference and our reasons for action.

But our account of the relationship between fitting motives and reasons to act offers a picture of how reasons for preference relate to reasons for action that is just as unified as TCR's assertion that all reasons to act are explained by our reasons to prefer their outcomes. WCP and MAP offer a systematic conceptual account of how the fittingness of motives explains the existence of reasons to act as they would motivate us. Because fitting intrinsic motives come in two varieties — state-directed and act-directed — this unified explanation entails that there will be intrinsic state-directed reasons corresponding to the former and intrinsic act-directed reasons corresponding to the latter. On this view, whenever

(RA) one has more reason to perform a_i than to perform a_j,

this is because

(FM) one has more fittingness reason to be motivated to perform a_i than to perform a_j.

In cases where all reasons to perform a_i and a_j are state-directed, (FM) will take the particular form

(FP) one has more fittingness reason to prefer o_i over o_j.

In cases where some reasons to perform a_i or a_j are act-directed, the instance of (FM) that explains (RA) will not take the form (FP), but it will explain (RA) just as well. Our explanation thus goes deeper than TCR's. By drawing on considerations of how practical reasoning governs motivation, intention, and action, it explains *why it is* that fitting state-directed preferences account for our practical reasons when they do. But this same explanation also explains why fitting state-directed motives do not — although their fitting act-directed counterparts do — account for our practical reasons in other cases.[40]

40. We think that this constitutes a direct response to the broad unificationist argument Portmore (2011, 81–82) gives in §3.4.4, "In defense of TCR on the whole". In §3.4.1–3 he gives a more specific unificationist argument. This argument first defends the biconditional that one has more reason to perform a_i than a_j iff one has more reason to prefer o_i to o_j. It then argues that TCR's explanation of the biconditional (that greater reasons to prefer o_i to o_j always explain greater reasons to perform a_i over a_j) is superior to its three possible rivals. First, it can't be that greater reasons to perform a_i over a_j always explain greater reasons to prefer o_i to o_j, because some cases (*e.g.* the mutual fund case) don't fit this pattern. Second, Portmore argues that it would be unacceptably disunified to hold that sometimes greater reasons to prefer o_i to o_j explain greater reasons to perform a_i over a_j but sometimes vice versa. Third, Portmore argues that he can't see how there could be some third factor that always simultaneously explains both greater reasons to perform a_i over a_j and greater reasons to prefer o_i over o_j.

We strongly suspect that Portmore's biconditional is false. If you could save five individuals by killing one (say, by pushing the one into the path of a trolley about to kill the five), we are inclined to think that you should more strongly prefer that the five live, more strongly prefer that the one is pushed, and thus more strongly prefer that you push him, even though you should not push him. It might seem strange at first to think that we should hope that we will act as we should not act. But it is actually a familiar phenomenon that we should hope that we will have motives that it is unfitting to have (*e.g.* unwarranted anger towards one if that is the only way to prevent an evil demon from killing five). If, as we have argued, what there is reason to do just is a matter of what it is fitting to be motivated to do, it should be no more surprising for it to be reasonable to hope that we will *do* what it is unreasonable to do (*e.g.* unreasonably kill one individual if that is the only way to save five). While it is plausible that we should intrinsically value our acting reasonably,

Some may think that TCR receives support from decision theory, or the theory of how it is rational to pursue given aims in light of given expectations about states of the world.[41] The basic question of decision theory is usually put as one of how to choose from a set of acts that can be represented as functions from possible states of the world to outcomes that the decision maker values, or *gambles* of the form $\{(S_1, O_1), (S_2, O_2), \ldots, (S_n, O_n)\}$, which associate each possible state of the world S_i with an outcome O_i that will obtain if the act is performed and S_i obtains. If the decision maker's preferences among gambles obey certain rationality constraints, they can be represented by a utility function that makes the strength of her preference for each gamble a function of the strength of her preference for its possible outcomes.[42] The basic idea of how to make rational decisions under conditions of risk is to perform the act (or pick from among the set of acts) with the highest expected utility, which is obtained by multiplying one's degree of preference for each of the act's possible outcomes by the probability that the outcome will obtain if one performs the act.[43]

Since decision theorizing is done with such heavily teleological language, one might be tempted to think that our basic principles for how to make rational decisions would have to be abandoned or substantially revised if TCR were rejected. But we think this is an illusion. One way to understand decision theory is as a theory of what it is reasonable to do in light of one's evidence, *assuming* that one's aims and expectations are reasonable. Another (perhaps more common) way is to understand it as a theory of what to do *given* one's aims and expectations if one is to retain a kind of minimal, instrumental rationality.[44] Either way, the theory should be entirely non-committal as to the kinds of aims it is reasonable to have, and in particular whether they are state-directed or act-directed.

Apart from common labelling conventions, we do not think that there is anything in the framework of standard decision theory that supports the assumption that the decision maker's ultimate aims, which are taken as given or assumed to be reasonable, must be state-directed as opposed to act-directed motives. What the framework does is understand acts as gambles $\{(S_1, D_1), (S_2, D_2), \ldots, (S_n, D_n)\}$ that associate each state of the world S_i with the motivationally relevant

the value of this, even from our own perspectives, should be absolutely trivial compared to that of someone's life (let alone four lives). To look more fondly upon one's following the dictates of reason than upon one (or certainly *four*) other individuals' surviving would be monstrously narcissistic. (The examples Portmore uses to support the biconditional appear ill-chosen; as he admits [115 n76], it is pre-theoretically plausible that one should more strongly prefer one's children living than strangers living, and the Kantian idea that one should not lie to a murderer to save his victims is distractingly absurd.)

Still, if Portmore's biconditional did turn out to be true, we believe our account would give, as he demands, a principled "explanation as to why sometimes reasons for acting are explanatorily prior and other times reasons for desiring are explanatorily prior" (2011, 81). Our explanation of how act-directed reasons for action explain the fittingness of pro-attitudes towards the state of our acting on them does rely upon the substantive view that it is fitting to prefer states in which one does what one has most reason to do. Because the proponent of TCR does not seem to need such an additional substantive view to explain Portmore's biconditional, there would be a case that TCR's explanation is simpler. But this simplicity would have to be weighed against TCR's costs. We have argued that there are good functionalist reasons to believe that we have act-directed as opposed to merely state-directed motives and good normative reasons to think that they are sometimes fitting. We have also argued that the very principles (WCP and MAP) that vindicate what truth there is in TCR's order of explanation entail that the fittingness of act-directed motives generates act-directed practical reasons. Moreover, there are serious worries about the narcissistic character of some agent-relative reasons when they are cast as state-directed. Against all this, we do not believe that the greater simplicity of TCR would be much of a benefit.

41. For instance, Pettit assumes that the plausibility of non-consequentialist views can be dispelled by the following decision-theoretic reasoning: "If one option has [better possible outcomes than all alternative options] such that it represents a better gamble than [those alternatives]…, then surely that is the best option for me to take" (1991, 239).

42. The main classic results are due to Ramsey 1926, von Neumann and Morgenstern 1944, and Savage 1954.

43. This principle has been put in teleological terms since its first clear formulation by Arnauld and Nicole: "In order to judge of what we ought to do in order to obtain a good and to avoid an evil, it is necessary to consider not only the good and evil in themselves, but also the probability of their happening and not happening, and to regard geometrically the proportion which all these things have, taken together" (1662, 367).

44. See Darwall 1983, Broome 1991, Gibbard 1998, and Joyce 1999.

description D_i that will be true of the act if S_i obtains. In cases where all one's intrinsic motives for or against performing the act are state-directed, these act-descriptions will include only the outcomes that the act will bring about if the state obtains. But in cases where one has intrinsic act-directed motives that might (depending on the state of the world) favor or disfavor the act, these descriptions will include properties of the act (like *is a keeping of a promise*, *is a killing of an innocent*) other than the outcomes it brings about. We can thus understand the standard results of decision theory as telling us that if the decision maker's choices among acts (*i.e.* gambles) obey certain rationality constraints, they can be represented with a "utility function" that represents the strength of her motivations to perform various acts, where the strength of her motivations to perform an act when its properties are uncertain is a function of the strengths of her motivations to perform it if its properties were known with certainty.[45] We think that this framework actually enables us to see more clearly the plausible idea embodied in the principle of maximizing expected utility, namely: the more averse you are (or should be) to performing an act with certain properties, the less willing you should be to perform an act if there is a risk that it will have those properties, and the greater the expectation you will need of its having properties that (should) incline you to perform it in order for it to be rational for you to take this risk. Since this is the principle that makes the standard theory of rational decision making plausible, and it applies just as easily to cases where some of the decision maker's ultimate aims are act-directed, we think that the plausibility of standard decision theory in no way counts against the existence of irreducibly act-directed practical reasons.[46]

45. Note that, under conditions of risk, the representations with which the agent's act-directed and state-directed motives will be combining to explain acts and act-directed motives will be *degrees of belief or credence* that acts will be parts of ways of performing other acts and that acts will bring about states of affairs. One therefore should not take our accounts of act-directed and state-directed motives in section 2 to be restricted to representations that are flat-out beliefs, or to the sorts of explanations of actions and act-directed motives that are produced by the motives combining with flat-out beliefs.

5. Practical Reasons to Be Moral, *Whether or Not* Morality Promotes the Good

We have thus argued that the same principles that explain why there are reasons to promote good outcomes, or act as fitting state-motives dictate, equally explain how there can be reasons to do other things, which fitting act-directed motives dictate. This is sufficient to defeat the logically strongest version of the top-down strategy of arguing for consequentialism, which holds, on the basis of TCR, that moral reasons must be reasons to promote good outcomes because all practical reasons are reasons to promote good outcomes. Our explanation of how reasons to act as fitting act-directed motives dictate are generated

46. We take this to be a slightly more detailed explanation of why, as Hurley (1997, 123–124) claims, act-directed goals are consistent with a maximizing conception of practical rationality. One could summarize the plausible general maximizing principle as instructing us to perform the acts that have the greatest expected degree of support by reasons (if we assume our aims are reasonable); or the greatest expectation of realizing the objects of our intrinsic motives, weighted by their strengths (if formulated as a principle that takes our aims as "given"). It is important to clarify that the maximizing principle so understood does not say that we should each have only one intrinsic super-aim to do something like "maximize the expected reason or motive support of my actions". An act's degree of expected reasons support is not some new thing that agents are supposed to care about over and above the fitting aims that it can be expected to achieve; it is simply a summary of how it is rational to trade off the act's expectations of achieving those fitting aims in light of their importance and its likelihood of achieving them.

    We take one of the main lessons of the consequentializing project of authors like Portmore to be that, if sufficient relativization is allowed in the descriptions of states of affairs, for any set of aims that includes intrinsic act-directed motives, there is a set of purely state-directed intrinsic aims that will motivate the same conduct in the same circumstances. The reason why it matters whether an ordering of gambles represents intrinsic motives that are act-directed or purely state-directed is that the state-directed motives may not be as fitting as the act-directed ones. For example, as we will suggest below, it seems plausible that it is fitting to feel more strongly obligated not to kill someone than to prevent five others from dying. But it seems unfittingly narcissistic to more strongly desire a state of affairs in which five die and you kill no one than a state in which only one dies and you kill someone.

in the same way as reasons to promote good outcomes should demystify how there could be practical reasons to do anything other than promote good outcomes.

But just because there can be practical reasons to act as fitting act-directed motives dictate doesn't mean that there are *moral* reasons to do so. Many act-directed reasons, like to run for the sake of running or to perform or create an artwork, seem to have little to do with morality. Our argument against TCR thus leaves open a version of the top-down strategy which maintains that, although there may be act-directed practical reasons, they cannot be moral reasons. A proponent of this top-down argument for consequentialism might hold that to be moral, a reason must be sufficiently selfless, disinterested, or impartial, and that the only practical reasons that have these properties are reasons to bring about impartially good outcomes, or states that it is fitting for all moral agents to have pro-attitudes towards.[47]

In this section we argue that this logically weaker version of the top-down strategy is also unsound. We show how the same kinds of considerations that support understanding a state's status as good in terms of the fittingness of pro-attitudes towards it support understanding an act's moral status in terms of the fittingness of feeling obligated to perform or avoid performing it. But, given the principles connecting fitting motives to practical reasons that we defended in section 3, this supports a direct connection between an act's moral status and reasons for or against performing it, which holds quite independently of whether the act promotes the good. Our account of moral reasons is logically consistent with the consequentialist view that all moral reasons are reasons to promote the good. Our primary aim in this section is simply to extend our argument against the top-down strategy to the logically weaker variety by showing that we do not need the idea of an act's promoting the impartial good to make sense of how there can be reasons to perform it that are both practical *and* distinctively moral. That said, we will show in the next section how our account of moral reasons can help defend the plausibility of non-consequentialist moral reasons, for instance against concerns about narcissism.

47. Although most consequentialists seem to embrace TCR, the idea that legitimately moral reasons must favor promoting those states we should prefer from an impartial perspective plays a central role in the arguments for consequentialism of such authors as R.M. Hare (1963, ch. 6–7), Singer (1979, 10–13), and Railton (1986, 189–190).

Just as there is a wide diversity of states that people can coherently judge to be good, there is a wide diversity of acts that people can coherently — if in many cases quite falsely — judge to be morally wrong or opposed by moral reasons. These include such plausibly wrongful acts as inflicting harms, defecting in collective action problems, and failing to respect autonomy, but also all manner of apparently miscellaneous acts, including sexual practices, drug use, swearing, "playing God", and so on (quite independent, in many cases, of their perceived consequences).

We think that the only thing that unifies the content of all these judgments is that those who make them think that we should feel obligated not to perform the acts in question. Feelings of obligation are, as Richard Brandt (1959, 117–118) observed, what you have when you see someone in trouble and feel like you "just can't" leave her. J.S. Mill (1863) described the feeling as an "internal sanction of duty … a feeling in our own mind … attendant on violation of duty, which in properly cultivated moral natures rises, in the more serious cases, into shrinking from it as an impossibility", and "a mass of feeling which must be broken through in order to do what violates our standard of right". The phenomenology of feeling obligated not to do something is similar to that of feeling guilt for having done it, but whereas guilt is retrospective, feeling obligated not to do something involves a kind of prospective guilt-tinged aversion to doing it.[48]

48. It is important, however, to emphasize that feeling obligated not to do something involves an aversion to *doing it*, not to the prospect of feeling guilt for having done it. If you saw someone in need of help but had on hand a pill that would prevent you from feeling guilt for failing to help her, your feeling that you "just can't" leave her (unlike an aversion to feeling guilt) would motivate you to help her and generate no motivation at all to take the pill.

    One might worry that feeling obligated not to do something involves judging that it would be wrong to do it, in which case it would be circular to

Thus, for reasons similar to those that favor analyzing judgments that states are good as judgments about the fittingness of pro-attitudes towards them, we think that the content and normative force of judgments that acts are wrong or opposed by moral reasons are best captured by analyzing them as judgments about the fittingness of feeling obligated not to perform them. For instance, what seems distinctive about viewing the fact that doing *A* will save someone's life as a *moral* reason to do *A* is one's taking this consideration to count in favor of feeling obligated to do *A*.[49] Similarly, what seems distinctive about thinking that the fact that doing *A* would kill someone makes it morally wrong or forbidden (as opposed to just unreasonable) to do *A* seems to be one's taking this consideration to make it, on balance, fitting for you to feel obligated not to do *A*.

try to explain judgments about wrongness in terms of judgments about the fittingness of feelings of obligation. But the fact that we can recalcitrantly feel obligated not to do things that we judge perfectly permissible seems to show that the feeling need not involve the judgment. For instance, a gay person raised in a homophobic culture might be fully confident that it is permissible to have sexual relationships with people of the same sex, but still have lingering feelings of obligation not to do so. (This is an instance of the general objection from recalcitrant emotions to the "judgmentalist" view that emotions involve ethical or evaluative judgments — see Gibbard 1990 and D'Arms and Jacobson 2003.)

49. The best alternative proposal about what is distinctive about viewing this as a moral reason is presumably that it involves one's taking it to be a reason that one has simply because one's act will promote the well-being of the individual in question. But it is surely *coherent* to think that there are distinctly moral reasons to do things other than promote well-being: with some plausibility one can think there are intrinsic moral reasons to respect autonomy and keep promises, and we know only too well what someone is thinking when she takes the alleged fact that an act is "unnatural", "against tradition", or "against God's will" as an intrinsic moral reason against doing it. Moreover, although many of us are decent enough to accept a substantive principle of beneficence according to which there is intrinsic moral reason to promote the well-being of every individual capable of well-being, it is, sadly, coherent to think otherwise. The view that there are individuals whose well-being there is no intrinsic moral reason to promote (although perhaps still some intrinsic non-moral reason to promote) has been coherently entertained, for instance, by some who take exalted views of the moral relevance of such factors as retribution, autonomy, promise-keeping, supernatural wills, and group-loyalty.

This supports the following analyses of our concepts of MORAL REASONS and MORAL WRONGNESS:

> **Fitting Attitude Analysis of Moral Reasons**: To judge that *R* is a moral reason for agent *X* to ϕ is to judge that *R* is a fittingness reason for *X* to feel obligated to ϕ, and
>
> **Fitting Attitude Analysis of Moral Wrongness**: To judge that it is morally wrong for *X* to ψ is to judge that it is, on balance, fitting for *X* to feel obligated not to ψ.[50]

50. Of course, we can think it perfectly fitting for someone to experience no feelings of obligation to refrain from doing things we think wrong if she is already sufficiently motivated not to do them. In most cases we would never even consider doing things that would kill others, and if we do, care for those others and fear of punishment are almost always sufficient deterrents. Although we think it would be wrong for us to kill in such cases, we surely do not think it inappropriate that we experience no feelings of obligation to refrain from doing so. Moreover, there is a sense in which we can think it fitting on balance to feel obligated to do things that we do not think it wrong to fail to do. It seems perfectly fitting for someone who goes above and beyond what morality requires — say, by getting killed to save a younger stranger from death — to feel obligated to do what she does.

    To clarify our proposal, it is important to note first that talk of feeling emotions, like talk of desiring or preferring, is ambiguous between an occurrent and a dispositional sense. Occurrent feelings and preferences exert causal pressure on one's behavior at the moment, and (at least typically) involve phenomenal experiences, while dispositional feelings and preferences merely have the disposition to become occurrent in certain circumstances. Thus one can dispositionally feel obligated not to push one's friends out of windows in the same way one can dispositionally feel anger at one's father even while one is enjoying his company and experiencing no negative emotions. Second, it is important to note that a response's being "fitting on balance" is ambiguous between (i) the response's being *mandatory*, in that there is no alternative response that is *as* strongly supported by fittingness reasons, or (ii) the response's being *justified*, in that there is no alternative response that is *more* strongly supported by fittingness reasons.

    In more detail, then, our proposal is that to think it morally wrong for *X* to ψ is to think that it is mandatory for *X* to have at least a dispositional feeling of obligation not to ψ (and mandatory for *X* to have an occurrent feeling of obligation not to ψ only if *X* is not already sufficiently motivated not to ψ). The sense in which one can judge it "fitting on balance" for *X* to feel obligated to ϕ when one takes *X*'s ϕ-ing to be supererogatory is that one thinks *X*'s feeling of obligation is justified but not mandatory.

Just as judgments about the goodness of states have the central normative feature of guiding pro-attitudes towards them, moral judgments seem to have the central normative property of guiding feelings of obligation. These fitting attitude analyses of moral judgments can explain their ability to generate motivation to act out of feelings of obligation as a special case of the ability of judgments that attitudes are fitting to directly guide us into having them.

To appreciate the centrality of this attitude-guiding role of moral judgments, suppose that someone were to label as "morally wrong" all those things we would call morally wrong, but took this to have no significance for what it was appropriate to feel obligated to do and consequently had no propensity to feel obligated not to perform the acts in question. It seems that by 'morally wrong' she would not really mean morally wrong. On the other hand, if someone were to label as "morally wrong" precisely those things we think permissible, she would still seem perfectly intelligible as thinking that those things are morally wrong so long as she thought it was fitting to feel obligated not to perform them. That said, as in the case of our fitting attitude analysis of good states, we will not rely on this strong claim that judgments about the fittingness of feelings of obligation *exhaust* the content of moral judgments. All we require is the claim that judgments about the existence of fittingness reasons for feelings of obligation are entailed by or *part* of the content of moral judgments.[51]

Given these fitting attitude analyses of moral concepts, the exact same connection between fitting motives and reasons to act that explains why there are reasons to promote the good equally explains a direct conceptual connection between an act's moral status and the existence of practical reasons for or against performing it. There is thus no need for the link between morality and practical reasons to be mediated by the link between good outcomes and practical reasons, and consequently no need for moral considerations to be reasons to promote the good in order for us to see how they can be genuine reasons to act. Since a judgment's truth entails the truth of its analysans, it is, according to our fitting attitude analyses of moral judgments, a conceptual truth that (i) *R is* a moral reason for *X* to ϕ iff *R is* a fittingness reason for *X* to feel obligated to ϕ, and (ii) *X*'s ψ-ing *is* morally wrong iff it *is* fitting for *X* to feel obligated not to ψ.[52] Since, as we noted, feeling obligated to ϕ essentially involves motivation to ϕ (and feeling obligated

51. This analysis of moral concepts is defended at greater length by Nye (2009). There have been previous proposals about how to analyze moral judgments as judgments about the fittingness of certain attitudes, most notably the proposal of Gibbard (1990, 44–45, 126–150). Gibbard began by analyzing judgments that *X*'s ψ-ing is morally *blameworthy* as judgments that it is fitting for *X* to feel guilt for ψ-ing and fitting for others to feel anger at *X* for ψ-ing. Gibbard then proposed that we analyze judgments that *X*'s ψ-ing is morally wrong as judgments that *X*'s ψ-ing would be morally blameworthy absent excuse. While we agree that there is a conceptual connection between wrongness and blameworthiness, Gibbard's 1990 analysis offers no way of interpreting conflicting normative judgments about excuses, and does not explain the role of normative judgments in guiding prospective behavior. We believe that adopting the analysis of moral wrongness defended by Nye (2009) together with Gibbard's analysis of moral blameworthiness can overcome these problems. In more recent work, Gibbard (2008) has proposed to do essentially the same thing.

52. As with the connection between good states and fitting pro-attitudes, what is most central to our argument is the *existence* of these connections between an act's moral status and fittingness reasons to feel obligated to perform or avoid performing it, rather than the particular explanation of this connection provided by our fitting attitude analyses of moral judgments. We do believe that our basic argument could be made to work with an alternative explanation of this connection, for instance that what makes a consideration a fittingness as opposed to a non-fittingness reason to feel obligated to ϕ is that it is in fact a moral reason to ϕ. We suspect, however, that such an explanation of the connection between morality and fitting attitudes might be more naturally paired with an explanation of MAP according to which an act's moral status jointly explains both (a) the fittingness of feelings of obligation to perform or avoid performing it, and (b) the existence of practical reasons to perform or avoid performing it. For reasons we observed in note 34, this purported explanation of MAP would seem to give opponents of our argument more room to hold that an act's moral status determines (a) without necessarily determining (b). Because we suspect that our fitting attitude analysis of acts' moral statuses is the explanation of the connection between morality and fitting attitudes that is most naturally paired with our explanation of MAP, we suspect that it will be the least vulnerable to opponents' attempts to argue that MAP holds in the case of fitting motives to bring about states but fails to hold in the case of fitting feelings of obligation to do things that may not bring about states towards which it is fitting to have pro-attitudes.

not to ψ essentially involves motivation not to ψ), it follows from (i) and (ii) together with the Warrant Composition Principle that (i′) if *R* is a moral reason to ϕ, then *R* is a fittingness reason to be motivated to ϕ, and (ii′) if ψ-ing is morally wrong, then it is fitting to be motivated not to ψ. Moreover, it follows from (i′) and (ii′) together with the Motivations-Actions Principle that (i*) if *R* is a moral reason to ϕ, then *R* is a genuine practical reason to ϕ, and (ii*) if ψ-ing is morally wrong, then there are genuine practical reasons not to ψ.

For instance, suppose (as seems overwhelmingly plausible) that the fact that an act will cause someone pain is an intrinsic moral reason not to perform it — which, absent factors like its preventing even more egregious harms to others, will make it wrong to perform it. Given the fitting attitude analyses of moral concepts, this entails that the fact that an act will cause someone pain is a reason to feel intrinsically obligated not to perform the act (*i. e.* to feel obligated not to perform the act simply on account of its being a causing of her pain) — which, absent other factors, will make it, on balance, fitting to have this feeling of obligation. Since feeling intrinsically obligated not to cause someone pain involves intrinsic motivation not to cause her pain, WCP entails that the fact that your act would cause someone pain is equally a fittingness reason to be intrinsically motivated not to perform the act. Finally, MAP entails that this fittingness reason to be intrinsically motivated not to cause her pain is actually an intrinsic practical reason not to cause her pain.

Thus, there is no need to think that a moral reason against doing something must be a reason against bringing about a sub-optimal outcome in order to see how it can be a genuine practical reason against doing it. The mere fact that the consideration counts morally against an act directly entails that it is a fittingness reason to feel obligated not to do it, which entails that it is a *practical* reason against doing it. There is no reason in the abstract why these reasons cannot be entirely act-directed: it is plausible to think that the fact that your act will cause Leslie pain is a reason to feel obligated *not to perform that act*, as opposed to simply a reason to feel obligated to bring it about that she is not in pain. Absent special obligations to either Leslie or Gertrude, there is no stronger reason to feel obligated to *bring it about* that Leslie is not in pain than to *bring it about* that Gertrude is not in pain.[53] But it is plausible that the fact that your act will cause Leslie pain is a stronger reason to feel obligated *not to perform it* than your reasons to feel obligated to bring it about that Gertrude is not in pain. If the only way to bring it about that Gertrude is spared equal or even slightly greater pain is to inflict significant pain on Leslie, it is plausible that you should still feel most strongly obligated not to inflict pain on Leslie. If these plausible thoughts are true, the connection between fitting attitudes and reasons to act entails that, all else held equal, there is stronger act-directed practical reason against inflicting pain on others than simply bringing it about that others are not in pain.

In fact, it follows from the fitting attitude analyses of moral concepts that, in a certain sense, *all* moral reasons are actually act-directed. All feelings of obligation are feelings of obligation to perform or omit *acts*, and essentially involve motivations *to perform or omit* them, so their fittingness always makes for practical reasons that are in some sense act-directed. It is true that one can feel obligated to perform an act *in virtue* of its having the property of bringing about a certain state of affairs. Plausibly, there are basic moral reasons of beneficence that count in favor of feeling obligated to perform acts simply in virtue of their bringing about the well-being of others. But these reasons are still *fundamentally act-directed*, in that they ultimately derive from the fittingness of attitudes towards acts, even though they are intrinsic reasons to bring about states simply because of what those states involve.[54]

53. Moreover, there does not seem to be any stronger intrinsic reason to feel obligated to bring it about that there is less pain-inflicting in the world, less pain-inflicting done by oneself, or less pain-inflicting done by oneself now. As we will explain below, these suggestions seem much more vulnerable to charges of objectionable narcissism than the idea that there is stronger intrinsic reason to feel obligated not *to inflict* pain.

54. In the same way, instrumental reasons to perform acts so as to bring about good states of affairs may be *fundamentally state-directed*, in that they ultimately derive from the fittingness of attitudes towards states, even though they are reasons to perform acts. Indeed, even intrinsic reasons to perform acts could

So *whatever* considerations turn out to be moral reasons, they are conceptually guaranteed to be fittingness reasons to feel obligated to perform or omit acts, and consequently genuine practical reasons to perform or omit them. As we have seen, these practical reasons generated by the fittingness of act-directed motives are just as basic as, and need no sanction from, the state-directed reasons generated by the fittingness of state-directed motives which constitute the goodness and badness of states of affairs. So even if we reject the consequentialist thesis that all moral reasons must be reasons to promote the good, it should be just as clear why there is practical reason to be moral as why there is practical reason to promote the good. This, we believe, should demystify not only how there can be genuine practical reason to do things other than promote good outcomes, but how some of these reasons can be genuinely moral.

6. Conclusion

We have thus argued that, because there is no mystery about how there could be reasons to do things other than promote the good that are both practical and moral, the top-down strategy of arguing for consequentialism fails. The deep theoretical connection between good states and reasons to promote them is a special instance of a more general connection between fitting attitudes and reasons to act as they would motivate us. Another instance of this connection is that between an act's moral status and reasons for or against performing it, which holds quite independently of whether morality directs us to promote the good. So once we understand *why* good states are connected to practical reasons, we can see that morality has the exact same kind of direct theoretical connection to practical reasons, and does not need to borrow its practical force from good states by prescribing that we promote them.

If our argument is sound, we believe that it undermines the case for preferring relativized versions of consequentialism like Portmore's to non-consequentialism, as this case relies almost exclusively on the teleological conception of practical reasons. Moreover, we think our explanation of how morality could give us entirely act-directed practical reasons helps us understand the strength of the case for preferring non-consequentialism to relativized consequentialism. The relativized consequentialist must hold that we are not allowed, for instance, to perform one murder to prevent five murders, because it is fitting for us to prefer the state of *our* not murdering to the state of four more lives being saved. While it seems appropriate to be somewhat specially concerned about one's own moral character, it seems monstrously narcissistic to look more fondly upon the state of not being a murderer oneself than on the survival of four individuals. The non-consequentialist, unlike the relativized consequentialist, can hold that it is fitting to value the survival of four individuals more than one's own moral purity, but that what states it is fitting to value doesn't always settle what to do. We are simply not permitted to save the five by performing a murder ourselves, because it is fitting to feel much more strongly obligated not to kill than to merely fail to save. This, we believe, is a much more plausible way of understanding agent-centered constraints.

As we indicated above, there is still a bottom-up case to be made against the initial plausibility of non-consequentialist ideas, so even if we are successful in dispatching the top-down strategy, impartial

be state-directed, if they were ultimately generated by the intrinsic desirability of the state of their being performed (this is essentially how proponents of relativized consequentialism portray reasons to observe agent-centered constraints). *R* is an intrinsic reason to respond in way *W* (*e.g.* to desire that Leslie is not in pain, bring it about that she is not in pain, feel obligated not to cause her pain, or omit causing her pain) if *R* counts in favor of *W* simply in virtue of describing what *W*'s object is in itself (*e.g.* Leslie's not being in pain or the act of causing her pain) and independent of this object's relation to anything else. But the "simply in virtue" here indicates only that there is no further *normative factor* that makes it the case that *R* counts in favor of *W*. It is consistent with *R*'s being an intrinsic reason to *W* that there is some further *conceptual explanation* of how *R* comes to count in favor of *W*, such as by *R*'s counting in the first instance in favor of feeling obligated to do whatever will bring about the state of Leslie's not being in pain, and (given WCP and MAP) *R*'s consequently counting in favor of bringing it about that she is not in pain.

consequentialism will not be left unmotivated.[55] But, as we explained, many arguments that seem largely bottom-up tacitly rely on the top-down strategy in demanding a special kind of justification for moral reasons that do not direct us to promote the good. As such, we think the elimination of the top-down strategy significantly weakens the overall case for consequentialism. But the most salutary effect of our argument on the case for impartial consequentialism might be to force its proponents to distill the potentially powerful bottom-up elements from the unsuccessful top-down ones, and to explain more clearly why we should think the plausibility of non-consequentialist ideas cannot survive reflective scrutiny.

Finally, we think that our argument has the potential to undermine theories other than direct consequentialism that rely on the idea that, in order to be practically relevant, morality must achieve ends that we have non-moral reasons to care about. Although indirect consequentialists do not think moral requirements must direct us to promote the good, they still take a "design stance" towards morality by holding that, in order to be justified, a system of moral requirements must promote the good more than any other system society could adopt.[56]

Non-redundant forms of contractualism take a similar design stance towards morality by holding that moral principles are justified by the fact that everyone would (at least under the right conditions) have strongest non-moral reasons to agree to their adoption.[57]

If we are right about morality's direct connection to practical reasons, it makes no more sense to take this pragmatic design stance towards what should count as morally right than it would to take it towards what should count as non-morally good or reasonable. We have argued that non-moral considerations are genuine practical reasons because they make for the fittingness of non-moral motives. But, as we have argued, moral reasons are considerations that make for the fittingness of feelings of obligation, which are guaranteed to be practical reasons by the same general connection between fitting motives and reasons to act. To paraphrase Frankena (1963, 98), it would give morality too despotic a role in our practical lives to insist that non-moral motives can be fitting only if they are "made for" or achieve what there is moral reason to achieve. But if we are right that the fittingness of moral motives is just as directly connected to practical reasons as that of our non-moral motives, it seems to give morality too servile a role to insist (as Frankena evidently did) that moral motives are fitting only if they are "made for" or achieve what there is non-moral reason to achieve.

55. These non-consequentialist ideas are exactly the ones relativized consequentialists seek to preserve by "consequentializing" them, so their being undermined would support impartial consequentialism over both non-consequentialism and relativized consequentialism. While Portmore (2011, 103–111) provides interesting "bottom-up" arguments against certain ("victim-focused") versions of non-consequentialism, he explicitly acknowledges that that these arguments do not work against all versions of non-consequentialism. Our point in the last paragraph suggests, moreover, that if you want to be "agent-focused" about agent-centered constraints, it is better to treat them as act-directed considerations that make it fitting for agents to be more motivated to omit certain acts than as state-directed considerations that make it fitting for agents to be more "concerned about" their performing them.

56. See for instance Brandt 1967 (esp. 114) and Hooker 2000 (esp. ch. 1).

57. See Harsanyi 1953; Brandt 1988; Gibbard 2008, parts II–III; and Parfit 2011, ch. 15–17. Of course, there are versions of contractualism which hold that moral principles are justified just in case no one could reasonably reject them, where moral considerations can determine what it is reasonable to reject. As McGinn 1999, Kamm 2002, and Hooker 2002 argue, Scanlon 1998 slips into such a view in deriving substantive moral conclusions from his version of contractualism, and Rawls 1971 (esp. sec. 4) seemed to slip into such a view in ecumenically motivating the veil of ignorance as an epistemic device for purging ourselves of biases. But, as McGinn, Kamm, Hooker, and Parfit note, by presupposing the moral reasons the device of agreement is supposed to explain, these versions of contractualism make the device redundant and suffer from vicious circularity.

    Because the design stances of indirect consequentialism and contractualism are so similar, it is not surprising that so many indirect consequentialists have supported their views by appealing to contractualist ideas. Indeed, considerations offered by Gibbard 2008 and Pogge 1995 strongly suggest that any genuinely non-redundant version of contractualism will be a form of indirect consequentialism.

References

Anderson, Elizabeth. 1993. *Value in Ethics and Economics*. Cambridge, Mass.: Harvard University Press.

Arnauld, Antoine and Pierre Nicole. 1662. *The Art of Thinking, Port-Royal Logic*. Trans. J. Dickoff and P. James, New York: Library of Liberal Arts, 1964.

Bennett, Jonathan. 1966. 'Whatever the Consequences'. *Analysis* 26: 83–102.

———. 1981. Morality and Consequences. In S. McMurrin (ed.), *The Tanner Lectures on Human Values*, vol. II, Salt Lake City: University of Utah Press, 47–116.

———. 1995. *The Act Itself*. Oxford: Oxford University Press.

Blanshard, Brand. 1961. *Reason and Goodness*. London: George Allen & Unwin.

Brandt, Richard B. 1959. *Ethical Theory*. Englewood Cliffs, N.J.: Prentice Hall.

———. 1967. Some Merits of One Form of Rule-Utilitarianism. *University of Colorado Studies in Philosophy* 39–65. Reprinted in R. Brandt, *Moraltiy, Utilitarianism, and Rights*, New York: Cambridge University Press, 111–36.

———. 1988. Fairness to Indirect Optimific Theories in Ethics. *Ethics* 98: 341–360.

Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.

Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

Cooper, Richard and Tim Shalllice. 2000. Contention Scheduling and the Control of Routine Activities. *Cognitive Neuropsychology* 17: 297–338.

Crisp, Rodger. 2000. Review of Value… and What Follows by Joel Kupperman. *Philosophy* 75: 458–462.

———. 2006. *Reasons and the Good*. Oxford: Oxford University Press.

Danielsson, Sven and Jonas Olson. 2007. Brentano and the Buck-Passers. *Mind* 116: 511–522.

D'Arms, Justin and Daniel Jacobson. 2003. The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism). *Philosophy: The Journal of the Royal Institute of Philosophy*, 52 (suppl.): 127–145.

Darwall, Stephen L. 1983. *Impartial Reason*. Ithaca, N.Y.: Cornell University Press.

———. 2002. *Welfare and Rational Care*. Princeton, N.J.: Princeton University Press.

———. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, Mass.: Harvard University Press.

Dreier, Jamie. 1993. Structures of Normative Theories. *The Monist* 76: 22–40.

Dretske, Fred. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, Mass.: MIT Press.

Ewing, A.C. 1939. A Suggested Non-Naturalistic Analysis of Good. *Mind* 48:1–22.

Foot, Philippa. 1959. Moral Beliefs. *Proceedings of the Aristotelian Society* New Series 59: 83–104.

———. 1985. Utilitarianism and the Virtues. *Mind* 94: 196–209.

Frankena, William K. 1963. *Ethics*. Englewood Cliffs, N.J.: Prentice Hall, Inc.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, Mass.: Harvard University Press.

———. 1998. Preference and Preferability. In C. Fehige and U. Wessels (eds.), *Preferences*. Berlin: Walter de Gruyter, 239–259.

———. 2003. *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.

———. 2008. *Reconciling Our Aims: In Search of Bases for Ethics*. New York: Oxford University Press.

Grafton, Scott T. and Antonia F. de C. Hamilton. 2007. Evidence for a Distributed Hierarchy of Action Representation in the Brain. *Human Movement Science* 26: 590–616.

Hare, R. M. 1963. *Freedom and Reason*. Oxford: Clarendon Press.

Harsanyi, John C. 1953. Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking. *Journal of Political Economy* 61: 434–435.

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

——. 2002. Contractualism, Spare Wheel, Aggregation. *Critical Review of International Social and Political Philosophy* 5: 53–76.

Hurka, Thomas. 1987. 'Good' and 'Good For'. *Mind* 96: 71–73.

Hurley, Paul. 1997. Agent-Centered Restrictions: Clearing the Air of Paradox. *Ethics* 108: 120–146.

Joyce, James M. 1999. *The Foundations of Causal Decision Theory*. Cambridge, U.K.: Cambridge University Press.

Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Clarendon Press.

Kamm, F.M. 2002. Owing, Justifying, and Rejecting. *Mind* 111: 323–354.

Kavka, Gregory S. 1983. The Toxin Puzzle. *Analysis* 43: 33–36.

Lang, Gerald. 2008. The Right Kind of Solution to the Wrong Kind of Reason Problem. *Utilitas* 20: 472–489.

Lewis, David. 1989. Dispositional Theories of Value. *Proceedings of the Aristotelian Society, Supplementary Volumes* 63: 113–137.

Louise, Jennie. 2004. Relativity of Value and the Consequentialist Umbrella. *The Philosophical Quarterly* 54: 518–536.

McGinn, Colin. 1999. Reasons and Unreasons. *The New Republic* 24: 34–38.

Mill, John Stuart. 1863. *Utilitarianism*. London: Parker, Son, and Bourn.

Nielsen, Kai. 1972. Against Moral Conservativism. *Ethics* 82: 219–231.

Norcross, Alastair. 2003. Killing and Letting Die. In R. Frey and C. Wellman (eds.), *A Companion to Applied Ethics*, Oxford: Blackwell.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Nye, Howard L.M. 2009. *Ethics, Fitting Attitudes, and Practical Reasons: A Theory of Normative Facts*. Ph.D. dissertation, University of Michigan.

Parfit, Derek. 1997. Equality and Priority. *Ratio* 10: 202–221.

——. 2001. Rationality and Reasons. In D. Egonsson, J. Josefsson, B. Pettersson, and T. Rønnow-Rasmussen (eds.), *Exploring Practical Philosophy: From Action to Values*, Aldershot: Ashgate, 17–39.

——. 2011. *On What Matters: Volume One*. Oxford: Oxford University Press.

Pettit, Philip. 1991. Consequentialism. In P. Singer (ed.), *A Companion to Ethics*, Oxford: Blackwell, 230–240.

Pogge, Thomas W. 1995. Three Problems with Contractarian-Consequentialist Ways of Assessing Social Institutions. *Social Philosophy and Policy* 12: 241–266.

Portmore, Douglas W. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford: Oxford University Press.

Rabinowicz, Wlodek and Toni Rønnow-Rasmussen. 2004. The Strike of the Demon: On Fitting Pro-attitudes and Value. *Ethics* 114: 391–423.

Railton, Peter. 1986. Moral Realism. *The Philosophical Review* 95: 163–207.

Ramsey, Frank P. 1926. Truth and Probability. In P. Gärdenfors and N. Sahlin (eds.), *Decision, Probability, and Utility: Selected Readings*, New York: Cambridge University Press, 1988, 19–47.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: Dover Publications, Inc.

Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, Mass: Harvard University Press.

Scheffler, Samuel. 1994. *The Rejection of Consequentialism: Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Oxford: Oxford University Press.

Schroeder, Mark. 2007. Teleology, Agent-Relative Value, and 'Good'. *Ethics* 117: 265–295.

——. 2012. The Ubiquity of State-Given Reasons. *Ethics* 122: 457–488.

Sen, Amartya. 1983. Evaluator Relativity and Consequential Evaluation. *Philosophy & Public Affairs* 12: 113–132.

Singer, Peter. 1979. *Practical Ethics*. Cambridge, U.K.: Cambridge University Press.

Sidgwick, Henry. 1907. *The Methods of Ethics*, Seventh Edition. Reprinted 1981, Indianapolis: Hackett Publishing Company.

Skorupski, John. 1999. *Ethical Explorations*. New York: Oxford University Press.

——. 2010. *The Domain of Reasons*. Oxford: Oxford University Press.

Smart, J.J.C. 1956. Extreme and Restricted Utilitarianism. *The Philosophical Quarterly* 6: 344–354.

Smith, Michael. 2003. Neutral and Relative Value after Moore. *Ethics* 113: 576–598.

Suikkanen, Jussi. 2009. Consequentialism, Constraints and the Good-Relative-to: A Reply to Mark Schroeder. *Journal of Ethics & Social Philosophy* (March) http://www.jesp.org.

Taurek, John M. 1977. Should the Numbers Count? *Philosophy & Public Affairs* 6: 293–316.

Temkin, Larry S. 2003. Egalitarianism Defended. *Ethics* 113: 764–782.

Thomson, Judith Jarvis. 1994. Goodness and Utilitarianism. *Proceedings and Addresses of the American Philosophical Association* 67: 5–21.

Velleman, J. David. 2002. Motivation by Ideal. *Philosophical Explorations* 5: 89–103.

von Neumann, John and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press.

Way, Jonathan. 2012. Transmission and the Wrong Kind of Reason. *Ethics* 122: 489–515.

Wilson, George and Samuel Shpall. 2012. Action. In E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), URL = <http://plato.stanford.edu/archives/sum2012/entries/action/>

Zimmerman, Michael J. 2011. Partiality and Intrinsic Value. *Mind* 120: 447–483.