Fittingness Objections to Consequentialism

Forthcoming in Christian Seidel (ed.), *Consequentialism: New Directions, New Problems?* Oxford University Press.

12 September 2017

Richard Yetter Chappell

University of York

Introduction

Traditional critics of consequentialism, from Bernard Williams to Michael Stocker, have objected to the apparent implications of (maximizing) consequentialism for moral agency. The consequentialist agent—an agent who has fully internalized the truth of consequentialism and has the attitudes and dispositions (if any) that would be appropriate given the truth of the theory—may not seem a plausible contender for being a morally appealing or virtuous agent. According to the familiar caricature, the consequentialist agent would be "cold and calculating", have "one thought too many" before acting, would regard others in an objectionably instrumental fashion—as mere "receptacles of value"—and be incapable of genuine friendship.[1] Consequentialist moral theorists, for their part, have largely dismissed such character-based objections as irrelevant to the truth of consequentialism properly understood as a *criterion of rightness* rather than a proposed *decision procedure*.[2]

New work in the foundations of ethics—extending the *fitting attitudes analysis of value* to yield a broader notion of normative fittingness as a (or perhaps even the) fundamental normative concept—provides us with the resources to clarify and renew the force of traditional character-based objections to consequentialism. According to these revamped *fittingness objections*, consequentialism is incompatible with plausible claims about which

attitudes are truly fitting. If a theory's implications regarding the fittingness facts are implausible, then this can be taken to cast doubt on the truth of the theory.

§1 explicates how traditional character-based objections can be understood as challenging the consequentialist conception of a morally fitting agent, and why such 'fittingness objections' are a challenge to consequentialism itself. §2 explains why I take consequentialism to have fittingness implications, and why standard consequentialist responses to character-based objections is inadequate. §3 explores Railton's 'sophisticated' consequentialist psychology, and argues that it, too, fails to address the problem. §4 introduces the notion of 'well-calibrated' dispositions, by investigating the question whether it's always rational to act on a disposition that it's rational to acquire. Finally, in §5, I draw on this conception of 'well-calibrated' dispositions to show how I think the consequentialist can successfully respond to a range of paradigmatic fittingness objections.

1. The Fittingness Objection

The consequentialist—and especially, the utilitarian—agent is sometimes presented, in caricature, as one who calculates expected utilities before each decision, who finds the needs of those before his eyes to be no more salient than those inaccessible and far away, and who is ready and willing to commit atrocities in the name of efficiency, without hesitation or regret.[3] Such an agent seems morally perverse, far from exemplifying the kind of *ideal moral character* one would expect to find in an agent who has internalized the true moral theory and has the kinds of attitudes and dispositions that are morally appropriate or *fitting*.

You may initially doubt that consequentialism has *any* implications, deleterious or otherwise, for fitting attitudes. Full discussion of this concern must wait until §2.1. To get clear on the basic idea in the meantime, the relevant concept may be grasped via the following pattern:

It's fitting to desire that which is good or *desirable*, to admire the admirable, believe what is (genuinely) credible, and so on. So, for example, if utilitarianism holds that what's good is just the welfare of sentient beings, then the fitting utilitarian agent is one who desires just the welfare of sentient beings. If such desires are shown to be not actually fitting, then that's just to say that utilitarianism is false: it makes mistaken claims about which things are desirable.

Of course, what's desirable (*fitting* to desire) may come apart from what it would be *optimal* or *desirable to desire*, just as what it's fitting to believe (based on the evidence) may differ from the beliefs that are optimal (given various practical incentives).[4] When these two kinds of assessment diverge, consequentialists will insist that what matters, practically speaking, is the promotion of value. So it may be that we should, in such cases, try to bring it about that we have optimal-but-unfitting attitudes. (Such an outcome is, after all, itself fitting to desire and to pursue.[5] It may be that we ought to try to acquire a belief that *p* even if we ought not to believe that *p*. It may be rational to act so as to bring about an irrational belief or other attitude. "*What should I believe?*" and "*What should I bring it about that I believe?*" are very different questions—one answered by norms of belief, and the other by norms of action.) But the practical primacy of value should not be taken to imply that questions of fittingness lack theoretical import. As we'll see, the consequentialist *needs* to answer such questions if they are to offer an adequate response to character-based objections to their view.

Fittingness objections work as follows. We begin with a sketch of an agent's psychology that seems to accurately represent a consequentialist outlook or perspective, and yet also seems intrinsically defective or unfitting from a moral point of view. If it's true both that (i) the described agent accurately represents a "fitting consequentialist agent"—an agent that has fully internalized the truth of consequentialism and has the attitudes and dispositions (if any) that would be appropriate (fitting to their objects) given the truth of the theory—and yet (ii)

the described agent is morally unfitting, then these premises together cast doubt on the truth of consequentialism.

Why is this? Analytically, if an agent that qualifies as "fitting" according to candidate moral theory X is actually morally unfitting, then X is not the true moral theory. If X were the true complete moral theory, then the X-fitting agent would ipso facto be the morally fitting agent. Further, if an incomplete theory X cannot be coherently supplemented in such a way as to yield plausible verdicts about fittingness, then that casts doubt on X's claim to even be *part* of the complete moral picture. If it is shown to be incompatible with plausible verdicts about fittingness, that would be grounds for thinking the theory simply false.

This is, I believe, a powerful (and underappreciated) line of argument. Just as we have intuitions about what the morally *right action* would be in various cases—intuitions which must be brought into reflective equilibrium with any moral theory we can ultimately accept—so too we have intuitions about fitting attitudes and character traits, or what a virtuous person would look like. If we can reasonably assign some default trust to these intuitions, then it's a strike against a moral theory if it violates them. And if the violation is severe enough, such considerations could well provide decisive grounds for rejection.

That is how we should understand character-based objections to moral theories. Next I argue that the standard consequentialist dismissal of these objections is unwarranted. The form of the objection is one that needs to be taken seriously.

2. Why two standard responses fail

2.1 Must consequentialism have fittingness implications?

One might question whether there's any such thing as "the fitting consequentialist agent". I sometimes introduce fittingness talk in terms of what's rationally warranted from the "point of view" of a moral theory, but you might well wonder whether moral theories are really the sorts of things that can have perspectives. If they're not, or if consequentialism in particular needn't have any implications regarding fittingness, then aren't fittingness objections to consequentialism unable to get off the ground?

I have two broad replies to this line of concern. My first (and more ambitious) response is to try to convince you that these ideas do all make sense. But I also have a more conciliatory backup option in case this fails.

What is the "perspective" of a moral theory? It's just an abstraction from the perspective of a rational agent who has fully *internalized* the theory, and whose psychology thus reflects, in isomorphism, the dictums of the theory—being attuned to just the considerations that the theory identifies as morally significant. I find it plausible that any normative claim has some corresponding specification in the psychology of the fitting agent.[6] It may even be (though I don't need this, and won't argue for it, here) that these implications for the fitting psychology are what ultimately give *content* to our various normative concepts and the claims that we make with them. For example, given the conceptual link between *goodness* and *desirability*, to claim that the happiness of sentient beings is *good* straightforwardly implies that it is *fitting to desire* that sentient beings be happy. Anyone who failed to have such a desire would clearly not count as having properly internalized the thesis that happiness is good. More controversially: it may be that the appropriateness of desiring the general happiness is what gives content to the claim that it's good. This link to fitting agential responses is what makes the normative claim significant to us as agents: it has implications for how we should be.

Insofar as any particular consequentialist view presupposes a theory of the good, it has implications for the fittingness of the corresponding desires, at least. But if other sorts of normative claims are practically significant at all, they too must presumably have some sort of agential (fittingness) implications. Presumably *right* actions, for example, are those that it's fitting for us to choose (or to intend). If a theory marks certain kinds of acts—lying, say—as inherently and absolutely wrong, this might be reflected in a fitting psychology by *refusing to even entertain* such acts as options.[7] Any genuine moral considerations should presumably find some traction in the psychology of the fitting agent, whereas bad reasons (or non-reasons) shouldn't.

If I'm right about all that, then any moral theory will ultimately have fittingness implications, even if they aren't explicit in canonical statements of the theory. So long as we can identify which attitudes and habits of thought inevitably follow from the proper internalization of a moral view, then we have a grasp on what the "fitting agent", according to that theory, looks like. And we can then assess whether this agent plausibly is morally fitting.

But suppose I'm wrong about all that. Suppose we can't, strictly speaking, infer any fittingness claims (besides perhaps the value – desirability link) from the core tenets of a moral theory. And suppose one were to endorse a conceptually "sparse" form of consequentialism which made no explicit claims about value, either, but just directly specified that agents morally ought to maximize happiness (or whatever). Would such a sparse view still be subject to fittingness objections?

I think it would. This is because we could still raise questions about whether these basic deontic claims are *compatible* with plausible claims about fittingness. Even if the theory has no positive fittingness implications, it surely has negative implications, as there are constraints on what fittingness claims can coherently be combined with various deontic

claims. For example, the deontic claim that we ought always to maximize happiness is clearly in tension with claims that it's fitting to value or desire things other than happiness, or that a fitting agent would never even think of lying, or that it's fitting to prefer to save your child's life over that of two strangers.

So there remains a real challenge here. Even the most conceptually sparse consequentialist must either argue that such fittingness claims are *incorrect*—that they misdescribe what attitudes and patterns of thought are truly warranted or fitting—or else argue that their sparse consequentialism can be *supplemented* or developed in such a way as to coherently combine plausible fittingness claims with their original deontic claim. (Which of these two is the most promising strategy in any given case will, of course, depend on the details. For example, I don't think it's true that lying should always be unthinkable. But it seems right to me that certain kinds of actions should not generally be "on our radar", and so I go on—later in the paper—to show how consequentialists might accommodate this.)

For ease of exposition, I will continue to speak of the "fitting consequentialist agent". But if you are not convinced that consequentialism has positive fittingness implications, feel free to read this as shorthand for "the most plausible supplementary view of the 'fitting agent' that is compatible with consequentialism."

2.2 Criteria of rightness vs. decision procedures

In response to character-based objections, consequentialists standardly distinguish between *criteria of rightness* and *decision procedures*.[8] Just because utilitarians hold that an act is right when it maximizes expected utility (say), it doesn't follow that they recommend actually trying to calculate utilities in your everyday life. Indeed, given that such constant calculation

would be predictably counterproductive (due to lack of time, misleading evidence, cognitive bias, setting bad precedents, etc.), utilitarians would strongly recommend against it![9]

All this is true enough, but beside the point. I agree that it's an open empirical question whether being morally fitting (whatever that turns out to involve) would bring about good results in our actual circumstances (just as it's an open empirical question whether being rational more generally has positive instrumental value). If it wouldn't have good results, then consequentialism may recommend against its own internalization—the possibility of such "self-effacingness" is a familiar (and unproblematic) feature of the view. Nothing I've said denies any of this: I haven't claimed that we necessarily ought to try to become fitting agents, come what may. Any consequentialist should agree that there are more important things than the quality of our characters, after all.

The objection is not to consequentialism's *recommendations*, but to its *implications*. The standard consequentialist response assumes that the only way that consequentialists can assess decision procedures (and psychological elements more generally) is in terms of their instrumental value, or whether they're worth inculcating. In the previous section, I argued that this is not so. There's *also* a fact of the matter as to what the 'fitting' consequentialist psychology would be, quite independently of what psychology consequentialism *recommends* (on grounds of utility) that we try to inculcate. But if the fitting consequentialist psychology can be shown to be not actually morally fitting, that would—as previously explained—pose a serious problem for the view.

So we can't just ignore decision-procedures and other psychological elements. And nor can we merely settle for identifying those which best promote value, and are thus *recommended* by consequentialism. As normative theorists, interested in whether or not consequentialism is a true moral theory, we must also investigate what kind of psychology would be a *fitting*

psychology to possess, were consequentialism true. We can then assess whether this fitting consequentialist psychology is plausibly morally fitting, and hence whether consequentialism itself remains an eligible moral theory.

In the following sections, I explore two very different strategies for constructing a non-defective consequentialist psychology in answer to this challenge. First I consider the Railtonian "sophisticated" psychology, with non-consequentialist desires. Then I explain and defend my preferred account, according to which critics are mistaken to assume that an agent with fitting utilitarian motivations would thereby conform to their caricature.

3. Sophisticated Consequentialism

3.1 Explication

Railton contrasts two kinds of hedonistic (or, more broadly, consequentialist) psychologies, which we may consider as candidate views of what's fitting: 'subjective' and 'sophisticated' psychologies.[10] The subjective hedonist is solely motivated by concern for his own happiness. However, the 'paradox of hedonism' suggests that such a person is likely to end up quite unhappy. Happiness may be better achieved by those who are motivated by other concerns. Railton thus introduces the sophisticated hedonist—let's call her 'Sophie'—who "aims to lead an objectively hedonistic life (that is, the happiest life available to [her] in the circumstances) and yet is not committed to subjective hedonism."[11] Sophie may thus possess and act on distinctively non-hedonistic motives—e.g., concern for others—if such desires are conducive to her living a happier life overall.

Once she has moved beyond subjective hedonism, and acquired a happy collection of non-hedonistic motivations, we may begin to wonder in what sense Sophie is still a "hedonist" at

all, rather than a whole-hearted pluralist. What sets Sophie apart, according to Railton, is that her psychology continues to be regulated by a *counterfactual condition* according to which, despite her various desires, she "would not act as [s]he does if it were not compatible with [her] leading an objectively hedonistic life."[12]

Whereas the subjective hedonist regulates her individual actions according to hedonistic norms, Sophie's hedonism instead regulates her desires and dispositions. So, for example, her pro-friendship disposition may lead Sophie to perform individual acts that reduce her happiness—e.g. answering her friend's distraught 3 a.m. call—but her 'hedonic monitor' is not triggered to intervene unless it becomes clear that the relationship *as a whole* is detrimental to her happiness, such that she would be better off in the long run with different desires and dispositions.

One may question how this regulative mechanism is supposed to work. In particular, we may wonder whether Sophie has an overriding desire *to possess hedonically fortunate dispositions*, that she will act upon (overriding her other, non-hedonistic desires) whenever she's in a position to do so. But such an agent may be better described as a simple maximizer of happiness-promoting dispositions, rather than a sophisticated maximizer of happiness!

Sophie's hedonism is better understood as manifested not in a *desire* at all, but rather a higher-order mechanism that serves to regulate her desires through some sub-personal causal process. The key difference is that this hedonistic mechanism, unlike a desire, never directly manifests itself in action. It is not *itself* a motivation that she may act on (though it may cause her to acquire some independently motivating hedonistic desires, insofar as these would cause her to live a happier life). Its control over her actions is instead wholly indirect: The hedonic monitor shapes Sophie's desires in hedonically fortunate ways, and then she acts on *those desires*, whatever they may be.

The 'sophisticated' psychology may thus be described in two parts: First, there is the agent's overarching "primary goal", which she may identify with during reflective moments, but which does not tend to directly motivate her actions. Instead, she is moved by the "secondary" desires and dispositions that are produced and regulated by a mechanism that is responsive to her primary goal.

3.2 Evaluation

Supposing that the psychology described in §3.1 is coherent, it's an interesting question how exactly we should evaluate it. Is it a plausible account of the fitting consequentialist psychology? According to (egostic) hedonism, one's own pleasure is the only end that's truly desirable, or worth pursuing. Sophie then seems irrational, by hedonistic lights, in that her desires are not necessarily directed at what is (according to this theory) desirable, and her actions likewise fail to be sensitive to hedonistic reasons: she often benefits others at her own expense. On the other hand, she is not *completely* insensitive to hedonistic reasons: Her desire-regulating faculty ensures that she maintains the desires that (the evidence suggests) it is hedonically best for her to have—and if circumstances change, so will her dispositions. This suggests an important sense in which Sophie's reflective hedonism is ultimately 'in control', even if it is not what moves her. We may thus need to draw a distinction between (local) act and (global) agent rationality, allowing us to say that *Sophie* is rationally fitting or responsive to reasons, even if her particular *actions* are not.

It's worth noting that even this vestige of rational sensitivity may, in special circumstances, make her worse off. Consider Parfit's example of the society of perfectly rational egoists, some of whom come to realize that it will advance their interests to become irrational in a specific respect: namely, if they become transparently disposed to follow through on their threats regardless of the costs to themselves.[13] Such a "threat-fulfiller" can then strap a bomb

to his chest, and threaten an egoist that he will detonate it (killing them both) unless the egoist complies with his whims. He can safely make such threats, because he knows the egoist would sooner comply than die. As Parfit further shows, the rational response for the remaining egoists is to turn themselves into transparent "threat ignorers", who are stably disposed to (irrationally) ignore threats no matter the costs to themselves. A threat-fulfiller will leave the ignorers alone, because he knows that if he were to threaten them, they would ignore him, and he would then detonate the bomb, killing them both. (Note that the threat-fulfiller will not *issue* threats that he expects will make him worse off. It is merely *fulfilling* threats that he does blindly.)

In comparison to the pure threat-ignorers, Sophie is more apt to have her instrumental rationality exploited. Given transparency, the threat-fulfiller will know that if he threatens Sophie, she will comply. For Sophie's regulating mechanisms will not allow her to maintain a disposition once it becomes clear that it is disastrous for her long-term happiness. And a threat-ignoring disposition becomes clearly disastrous as soon one is actually issued with a credible apocalyptic threat. So, a threat-fulfiller will know that he can safely threaten Sophie, and she will (if necessary change her dispositions and) comply rather than die. To avoid such exploitation, Sophie would have to alter her psychology so that she would become a *pure* (unregulated, insensitive) threat-ignorer—at which point she would no longer be a sophisticated hedonist. She would just be (however fortunately) irrational, by hedonistic lights.

We thus find that a Railtonian sophisticated psychology is by no means guaranteed to endorse itself as the most fortunate psychology to possess in every possible situation. But it offers a suggestive alternative to the standard conception of an instrumentally rational psychology. Insofar as we are drawn to the idea that rationality should not normally be a curse (even if it

may be in certain special circumstances), we may see Sophie's two-level psychology—with its capacity for her primary goal to control and regulate her secondary, action-guiding motivations—as an improvement over the subjective hedonist's unitary motivational structure. While acknowledging that Sophie's actions are often locally irrational (by hedonistic lights), we may be more concerned to evaluate her global rationality as an agent. In this respect, at least, she may at first glance seem more reasonable.

I think there are important grounds for doubting this conclusion, however. Let's return our attention from hedonism to utilitarianism. The sophisticated utilitarian—call her 'Sophu'—will have whatever motivations are most conducive to promoting the general welfare. So, in particular, if an evil demon threatens to torture an innocent population unless Sophu comes to intrinsically *want* them to suffer,[14] then Sophu will be led to acquire this fortunate but malicious motivation.[15] This is a good outcome, in the circumstances, as it prevents a lot of suffering. But if any desire is unfitting by utilitarian lights (or common intuition, for that matter), it is surely an intrinsic desire that others suffer. Sophu has, quite virtuously, made herself vicious. And note that it is not just her *actions*, but her *desires*—her very *self*, we might think—that is impugned here. She (non-instrumentally) desires what is blatantly undesirable, thus violating the most basic criteria for qualifying as a fitting agent. Yet she is still a sophisticated utilitarian. So this 'sophisticated' consequentialist psychology is not adequate as an account of the *fitting* consequentialist psychology.

The advocate of sophisticated utilitarianism might at this point defend Sophu's utilitarian credentials by pointing out that her *deepest commitments* remain pure and altruistic, even as they respond to the unfortunate circumstances by shaping her motivations in this malicious-but-instrumentally-valuable direction. So there at least remains *something* fitting about Sophu's psychology. But it nonetheless contains potential moral defects of a sort that cast

doubt on claims that she qualifies as a fitting utilitarian agent *overall*. So it is worth investigating whether we can do better with a more direct approach.

4. Rational Transmission and Well-Calibrated Dispositions

We can identify where the 'sophisticated' psychology goes wrong (or fails to accurately represent a fitting consequentialist perspective), by considering the relation between (i) the rationality of acquiring and maintaining a desire or disposition, and (ii) the rationality of 'acting on' the disposition, i.e. performing an action that the disposition characteristically disposes you towards. Consider the following simple principle of rational transmission:

> **(RT-past)** For any disposition D and act A that is characteristic of D: *If it was rational to acquire D then it is rational to perform A.*

Parfit's above-described case of the threat-fulfillers casts doubt on this principle. It may well be rational for a self-interested agent to acquire the threat-fulfilling disposition, but if (through some irrational quirk) a threatened target unexpectedly ignores the agent's apocalyptic threat, it is surely *not* rational for the agent to follow through and blow themselves up. Such disastrous stubbornness would seem, on the contrary, quite crazy.

Gauthier is not wholly convinced by this counterexample to RT-past, but suggests and endorses a weaker transmission principle, which we may formulate as follows for any disposition D and act A that is characteristic of D:[16]

> **(RT-present)** If it was rational to acquire D *and is rational to maintain it presently*, then it is rational to perform A.

This principle, if true, could potentially vindicate the rationality of the sophisticated consequentialist's actions (and hence, arguably, the fittingness of their motivating desires). However, Parfit points out that even this weakened transmission principle is susceptible to counterexamples, such as:

> *Schelling's Case.* A robber threatens that, unless I unlock my safe and give him all my money, he will start to kill my children. It would be irrational for me to ignore this robber's threat. But even if I gave in to his threat, there is a risk that he will kill us all, to reduce his chance of being caught. [... I]t would be rational for me to take a drug that would make me [transparently] very irrational. The robber would then see that it was pointless to threaten me; and since he could not commit his crime, and I would not be capable of calling the police, he would also be less likely to kill either me or my children. [...] But while I am in my drug-induced state, and before the robber leaves, I act in damaging and self-defeating ways. I beat my children because I love them. I burn my manuscripts because I want to preserve them.[17]

Parfit stipulates that these destructive acts are not necessary to convince the robber that you are irrational. So they have no good effects, though they stem from a disposition (namely, the disposition to act irrationally) that it is worthwhile, for extrinsic reasons, to acquire and maintain. Are these acts rational? I share Parfit's sense that they are not. So the transmission principles considered thus far fail, suggesting a robust disconnect between the rationality of acquiring and maintaining a disposition vs. the rationality of acting upon it.

The fundamental explanation for this disconnect is that an agent's dispositions can have other consequences besides producing downstream acts in the agent herself. In particular, you might be harmed or rewarded directly on the basis of whether you possess some disposition, independently of whether you act on it. This suggests that we can distinguish (i) dispositions that have high expected value, all things considered, and (ii) dispositions that have high expected value *in respect of the downstream actions they'll tend to produce*. We can call the

former class of dispositions 'desirable', and the latter 'well-calibrated'. Dispositions that are desirable but *not* well-calibrated we may call 'extrinsically desirable'. It is these extrinsically desirable dispositions that feature in Parfit's cases of 'rational irrationality', i.e. whereby it is rational to acquire and maintain such a disposition, but irrational to act upon it.[18]

While acknowledging this possibility, we may still think that there must be *some* transmission principles according to which the rational status of a general rule or disposition can be inherited by the particular acts it prescribes. And, indeed, the distinction I've just highlighted suggests an obvious candidate principle: we just need to restrict the dispositions in question to those that are 'well-calibrated', i.e. desirable for their (expected) impact on your downstream actions, rather than for other reasons. Consider the following transmission principle:

> **(RT-Calibrated)** For any dispositional set D and act A that is characteristic of D: *If D is* well-calibrated*, i.e. expectably good to possess in virtue of the downstream actions it tends to produce, then it is rational to perform A.*

This seems much more promising, though I discuss a residual concern in a note.[19]

In sum: While I am uncertain that any such transmission principle is ultimately vindicated, formulations that focus on the subset of dispositions that are *well-calibrated*, in my described sense, would seem to have the best shot. And, as we will see, these are just the dispositions that may be possessed by the 'subjective' act consequentialist agent, in contrast to the unfitting but (extrinsically) desirable dispositions that we saw could be part of the 'sophisticated' consequentialist psychology. We are now in a position to spell out what I take the fitting act consequentialist agent to look like.

5. The Act Consequentialist Agent

Suppose we accept my earlier suspicion that 'sophisticated' psychologies, with their extrinsically desirable dispositions, are not fitting consequentialist psychologies. The remaining option for defending consequentialism against fittingness objections is to spell out a non-defective 'subjective' consequentialist psychology. In attempting this task, I will especially make use of the idea that our account of the fitting consequentialist agent, while restricted to consequentialist motivations, may at least appeal to *well-calibrated*, if not merely extrinsically desirable, guiding dispositions. This restriction is one of the main features that sets apart my straightforward account of the fitting consequentialist psychology from the 'sophisticated' view explored in §3.

5.1 Motivating vs. Guiding Dispositions

Let's begin by distinguishing what I'll call 'guiding' and 'motivating' dispositions.[20] Our non-instrumental desires or *motivations* are our driving concerns, or what move us to action. They represent the goals we hope to realize through acting. On the other hand, this motivational 'oomph' can be steered or *guided* by strategies and heuristic dispositions that shape our behavioural responses in pursuit of those goals. We may think of our guiding dispositions as, roughly, the psychological manifestation of instrumental rationality. They take our desires as inputs, and output a suitable action or intention.[21]

The standard caricature of the consequentialist agent assumes that we can "read off" both kinds of dispositions from the moral theory. From its theory of the good—say, the utilitarian view that what matters is just the welfare of sentient beings—we get the fitting utilitarian motivations. That much I agree with: the fitting utilitarian will desire the welfare of sentient beings.[22] But the standard caricature also takes the 'maximizing' aspect of consequentialism

to settle the *guiding* dispositions of the fitting consequentialist agent: they will (allegedly) decide how to act by, in each instance, conducting an expected-value calculation, and then perform whatever action they judge to have the highest expected value. It is this feature of the imagined consequentialist agent that is responsible for so much of their apparent defectiveness (as we will see in §5.3). And it is this feature that I deny we should attribute to the fitting consequentialist agent.

Can we coherently reject the critic's assumption that the fitting consequentialist must have these defective guiding dispositions? I think we can. Note, first of all, that our primary example of a theory's fittingness implications (namely, the goodness – fitting desire link) concerns *motivating* dispositions. There's no such clear link between our moral theories and any putative implications for the fittingness of our *guiding* dispositions. So it seems prima facie open to us to dispute this assumption.

Moreover, I think there are strong theoretical grounds for expecting the assumption to be false. Namely: (standard maximizing) consequentialism is naturally understood as a view that simply combines *ordinary instrumental rationality* with a specification of the moral ends to be pursued.[23] This theoretical understanding strongly suggests that the fitting guiding dispositions should be determined by our independent account of instrumental rationality (as I will go on to explore). The distinctive positive work of consequentialism as a moral theory is just to add the theory of the good, with its associated implications for fitting motivating dispositions. The remaining distinctive content of consequentialism is negative: it simply denies that various *other* normative elements, e.g. deontological side-constraints, play any fundamental role in determining what's right (or what fitting moral reasoning is sensitive to).

Finally, I hope to offer an "existence proof" of the coherence of a fitting consequentialist psychology that lacks defective guiding dispositions. In particular, I'll show that utilitarian

motivating dispositions can be coherently combined with a plausible account of the instrumentally rational guiding dispositions. If there's some reason the resulting agent fails to qualify as a "fitting utilitarian", I think the onus is on the critic to explain why this is so.

To fulfil this task, I begin with a brief sketch of some 'well-calibrated' guiding dispositions which I take to be (a) prerequisites for competent human agency, and hence (b) constitutive of instrumental rationality, at least for agents with human-sized minds. I will then show how an agent with fitting utilitarian motivations could also possess these well-calibrated guiding dispositions. Since these guiding dispositions appear to be compatible with utilitarian (or other consequentialist) motivating dispositions, I conclude that there is no barrier to the consequentialist supplementing her theory in just this way to secure a plausible vision of the "fitting consequentialist" agent. I will wrap up by illustrating how my well-calibrated fitting consequentialist can be used to address prominent character-based objections to consequentialism.

5.2 Defective Deliberation and the Well-Calibrated Agent

Let's consider four central features of the fitting agent's guiding dispositions. Firstly—as perhaps the most obvious prerequisite for competent agency—we have *epistemic rationality*: that is, the agent must have well-calibrated expectations about their environment, or a basic understanding of what counts as evidence for what. They cannot take the roar of a dangerous predator as evidence that a cute puppy awaits them outside. They need to have generally reasonable beliefs about their environment, and about what would be effective means for realizing their ends (whatever those might be).

Next, at the borderline of the epistemic and the practical, we will find constraints on how the agent is disposed to allocate their limited *attentional* resources. They must be generally

attentive to possible threats and opportunities in their immediate environment, while also—in a calm moment, when appropriate—considering more abstract mental models of past and possible future scenarios (for sake of planning, self-evaluation, etc.). The details aren't too crucial for my purposes, but as we'll see, it's important that the fitting agent not dwell excessively on the past.

Third, the competent agent requires well-calibrated habits, instincts, or sub-personal "predispositions"[24]—an "auto-pilot" set, e.g., to avoid pain, be cooperative, and help others in need—to secure effective automatic behaviour in normal circumstances. One reason for this is that in time-critical situations, the agent cannot afford to pause to reflect on their situation at all. Often, a competent agent will be moved immediately (without conscious deliberation) to act, upon registering pertinent information about their environment. This is no mere behavioural reflex, as the agent is genuinely acting for reasons. But the rational processing goes on "below the surface".

Once equipped with such well-calibrated predispositions, the fitting agent may act on them without need for excessive self-monitoring or executive control, and—in so doing—they may trust that they are acting for the best. Our fitting agent may, in this way, reap the practical benefits of 'satisficing' without the theoretical baggage.[25]

The fourth and final component that I'll discuss here is the possession of well-calibrated *triggers* for executive oversight. On pain of regress, we cannot always deliberate about whether to start deliberating. So, as previously noted, the agent's *default* guidance must be from non-deliberative "predispositions". But when these are not up to the task—when, say, the agent is faced with novel or complex circumstances for which their predispositions aren't so well calibrated to deal with—the agent's sub-personal mechanisms must recognize this and respond by triggering explicit deliberation on the part of the agent.

In summary: the fitting human-like agent—if they are to be capable of acting competently in a wide range of 'normal' circumstances—will rely heavily on well-calibrated predispositions, rather than explicit deliberation or calculation, to guide their actions in pursuit of whatever their goals may be. That's just what it is for agents with human-sized minds to be instrumentally rational or have fitting guiding dispositions. And this will be so even if their goal is to promote the well-being of sentient creatures as much as they are able. This, I propose, is how we should understand the fitting consequentialist agent. They may have straightforwardly utilitarian (or whatever) desires, which are then translated into action via the above-described 'well-calibrated' guiding dispositions.

5.3 Addressing the objections

We are now in a position to assess how my conception of the fitting consequentialist agent stands up to various anti-consequentialist objections.

We can first note that my 'well-calibrated' fitting consequentialist will not be "constantly calculating". Absent any triggering of their executive faculty, the fitting utilitarian (for example) will respond *directly* to the salient needs of others—a child drowning in a pond, say—without mediation by explicit deliberation, let alone abstract judgments of "permissibility". In this way, the fitting consequentialist will not exhibit what Williams famously called "one thought too many".[26]

The fitting consequentialist's reliance on generally-reliable predispositions also undermines the objection that they would engage in "marginally-beneficial rule-breaking", such as breaking a promise whenever the benefits from doing so seem to even slightly outweigh the costs.[27]

Because overt calculation often goes awry, the competent consequentialist will—as we've seen—rely heavily on her generally reliable predispositions in everyday life, only pausing to reflect when her well-calibrated sub-personal mechanisms alert her to the need (say due to complex novel circumstances that her "auto-pilot" wasn't designed to deal with). Everyday promise-keeping is not exactly novel, so for the fitting agent the question whether to keep a promise *shouldn't even arise*, unless there's something special about the situation that calls for her executive oversight.

That's enough to defeat the claim that the fitting consequentialist would commonly engage in marginally-beneficial rule-breaking. But we may draw an even stronger conclusion. For suppose that our agent's executive oversight happens to be triggered. In a typical case, what should she conclude? We can stipulate that in fact the outcome would be marginally better if she broke her promise, but presumably the agent herself will not have any easy way of knowing this. (Among other things, she'd need to first consider the possibility of self-serving bias corrupting her judgment, and also to weigh the apparent benefits of rule-breaking in this instance against the long-run value of retaining a reputation for trustworthiness.) Maybe if she heard the booming voice of God reassuring her of this fact, then she could rationally go ahead and break her promise without further worry. Such behaviour no longer seems intuitively troubling to me. But in ordinary circumstances, when rule-breaking might seem more worrying, it's almost never going to be *clear* that rule-breaking is beneficial unless it is significantly (not merely marginally) so.[28]

So our agent is faced with an immediate choice: she can (i) break the rule even though it's not yet clear to her whether this would have good results on net; (ii) sink further cognitive resources into investigating a question that she probably shouldn't have bothered to ask in the

first place; or (iii) simply keep her promise and turn her attention to more important matters. It seems pretty clear that, in this sort of case, option (iii) is the way to go.[29]

In sum: Breaking a rule will generally only be obviously worthwhile in cases where it is also of significant benefit (in which case many would approve of rule-breaking anyway). If it's only of marginal benefit, this fact typically won't be sufficiently clear for a reasonably self-doubting, fallible agent to immediately act upon. And the low potential payoff means that it isn't really worth inquiring further: better just to stick with the generally-reliable rule of thumb. So a fitting consequentialist generally won't be found engaging in marginally beneficial rule-breaking after all. (They'd even share our intuition that there's something awfully dubious about any agent who would act that way.) This gives them the kind of stable predictability needed for others to regard them as eligible and (more or less) trustworthy partners for social cooperation.

This discussion brings out the fact that the standard caricature of a consequentialist agent assumes that they will be unreasonably overconfident in their ability to calculate expected values accurately. But even if a consequentialist *initially* judges (just based on the first order evidence) that they would do best to break some generally beneficial rule, they may also realize that most people who make such judgments in similar situations are mistaken. Since they have no particular reason to think that they are one of the lucky few who make this judgment correctly, the general fact serves as a kind of higher-order evidence that their initial judgment was mistaken. All things considered, then, a reasonable expected value judgment should, in this sort of circumstance, end up reinforcing the general rule rather than licensing typically-misguided unilateral rule-breaking.[30]

The objections considered thus far—that the consequentialist would have "one thought too many", and that they would engage in "marginally-beneficial rule-breaking"—suggest the

need to distinguish (i) the appropriate answer to a question, and (ii) whether a well-functioning agent would ask that question in the first place. The need for this distinction becomes especially apparent when we consider the following objection from Michael Stocker:

> Maximizers hold that the absence of any attainable good is, as such, bad, and that a life that lacks such a good is therefore lacking. I disagree. One central reason for my disagreement stems from the moral psychological import of regretting the absence or lack of any and every attainable good. This regret is a central characterizing feature of narcissistic, grandiose, and other defective selves. It is also characteristic of those who are too hard on themselves, who are too driven and too perfectionistic.[31]

This objection strikes me as deeply misguided. It may be unfortunate, and indeed even inappropriate ("defective"), to actively regret every little regrettable thing. But those things may be regrettable all the same. Crucially, this is not to say that a rational agent must regret them. It is more like a hypothetical imperative: *if* you closely attend to the features in question, this should induce feelings of regret. But it may be a kind of rational defect to attend to the wrong things, if there are more pressing matters to attend to. As we saw in §5.2, the fitting agent would allocate their attentional resources in a way that avoids excessive dwelling on hypotheticals. So we can agree with Stocker that the agents he describes are defective, without thinking that the maximizing consequentialist would exhibit any such trait. On my picture, the consequentialist will have only a conditional disposition to regret the lack of a good *insofar as she attends to this lack*. But she'll usually have more important things to attend to, so she shouldn't actually end up actively regretting things very often at all. She is, in this sense, appropriately respons*ive* to reasons for regret, without having to be constantly respond*ing* to them.

I've now shown how the well-calibrated fitting consequentialist avoids three of the 'character-based objections' extant in the literature. Equipping the consequentialist agent with well-calibrated guiding dispositions helps to undermine claims that the fitting consequentialist psychology is inherently defective.

5.4 Act vs. Rule Consequentialist Agents

In light of my appeal to rules and dispositions, some readers may be puzzled by my labelling the resulting agent a fitting 'act consequentialist' agent. To avoid any confusion on this front, let me wrap up by briefly characterizing what I take to be the two main differences between (fitting) act and rule consequentialist psychologies.

First, while both make use of rules, they do so in very different ways. The act consequentialist adopts 'rules of thumb' for *instrumental* purposes, but their fundamental aim (reflected in my account of the fitting motivating dispositions) makes no essential reference to rules: they just want to bring about the best possible outcome, and refraining from deliberation is one (guiding) strategy they might employ, at appropriate times, as a means to this end. Rule Consequentialism, by contrast, builds reference to rules into its criterion of right action, and hence the corresponding 'fitting psychology' must likewise accord some fundamental, non-instrumental significance to rules—e.g. in the agent's fundamental desires or motivating dispositions. (This then opens them up to distinctively characterological objections of 'rule-worship'.)

A second, more straightforward difference is that they may employ rules with very different contents. I've suggested that a fitting act consequentialist could (whilst retaining their fitting character) only make use of 'well-calibrated' dispositions—dispositions whose value stems from the improved quality of the actions they dispose the agent towards. But insofar as rule

consequentialism appeals to rules that it would be good to internalize *for whatever reason*, they may well end up calling 'fitting' even dispositions that are merely extrinsically desirable. In other words, the fitting rule consequentialist agent would look much more like the kind of 'sophisticated' agent described in §3.

For example, suppose that suffering is always bad, but that widespread adoption of retributive attitudes towards punishment would form part of the optimal moral 'code' in a certain society. Since rule consequentialism assigns direct, non-instrumental significance to the rules of the optimal code, the fitting rule-consequentialist agent would presumably have to have corresponding non-instrumental desires. That is, rule consequentialism, in the imagined circumstances, has the implication that retributive punishment is fitting to desire. But we've supposed that in fact it isn't desirable: suffering is *always* bad. (The fitting act consequentialist, by contrast, may only desire to promote the good. If better results would be obtained with different attitudes, then this may lead them to try to transform their character so that they no longer qualify as a fitting act consequentialist at all. But while their view might thus *recommend* retributive or other attitudes, it maintains the distinction between the practically recommended attitudes and the ones that are fitting to their objects as a matter of principle.)

This then provides the basis for a simple new argument against rule consequentialism. Rule consequentialism implies that literally *anything* (from retributive punishment to gratuitous torture) could be rendered fitting to desire, just by tweaking the incentives surrounding the creation and maintenance of the public's moral code. But these things are not, in such circumstances, fitting to desire. Such desires are supported by the "wrong kind of reasons"—it is not the objects of these desires that are desirable, but rather the state of possessing the desire itself (or something to do with promulgating the moral rule, depending on what exactly

is responsible for generating the good consequences). Rule consequentialism thus has false implications about what's fitting, and is thereby shown to be false itself.[32]

References

Anscombe, G. E. M. “Modern Moral Philosophy.” *Philosophy* 33 (1958): 1–19.

Bales, R. E. “Act-utilitarianism: account of right-making characteristics or decision-making procedures?” *American Philosophical Quarterly* 8 (1971): 257–65.

Chappell, Richard Yetter. “Fittingness: The Sole Normative Primitive.” *Philosophical Quarterly* 62 (2012): 684–704.

—. “Value Receptacles.” *Noûs* 49 (2015): 322–332.

Gauthier, David. “Rationality and the Rational Aim.” In *Reading Parfit*, edited by Jonathan Dancy. Oxford: Blackwell, 1997.

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

Kahneman, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.

Kapur, Neera Badhwar. “Why it is wrong to be always guided by the best: Consequentialism and friendship.” *Ethics* 101 (1991): 483–504.

Mackie, J. L. “Rights, Utility, and Universalization.” In *Utility and Rights*, edited by R. G. Frey. Oxford: Basil Blackwell, 1985.

Mason, Elinor. “Do Consequentialists Have One Thought Too Many?” *Ethical Theory and Moral Practice* 2 (1999): 243–261.

Mill, J. S. *Utilitarianism*. 1863. http://www.utilitarianism.com/mill2.htm [accessed 20/07/2015]

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

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

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

Pettit, Philip and Geoffrey Brennan. "Restrictive consequentialism." *Australasian Journal of Philosophy* 64 (1986): 438–455.

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

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

Regan, Tom. *The Case for Animal Rights*. University of California Press, 1994.

Scheffler, Samuel. 'Agent-centred restrictions, rationality, and the virtues.' *Mind* 94 (1985): 409–419.

Slote, Michael and Philip Pettit. "Satisficing Consequentialism." *Proceedings of the Aristotelian Society, Supplementary Volumes* 58 (1984): 139–176.

Stocker, Michael. *Plural and Conflicting Values*. Oxford: Oxford University Press, 1989.

Williams, Bernard. "A Critique of Utilitarianism." In J.J.C. Smart and Bernard Williams (eds.), *Utilitarianism: For and Against*. Cambridge: Cambridge University Press, 1973.

—. "Persons, Character and Morality." In *Moral Luck: Philosophical Papers, 1973-1980*. Cambridge: Cambridge University Press, 1981.

---

[1] See, e.g., Bernard Williams, "A Critique of Utilitarianism," in J.J.C. Smart and Bernard Williams (eds.), *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973); Bernard Williams, "Persons, Character and Morality," in *Moral Luck: Philosophical Papers, 1973-1980* (Cambridge: Cambridge University Press, 1981), 18; Tom Regan, *The Case for Animal Rights* (University of California Press, 1994), 208-211; Neera Badhwar Kapur, "Why it is wrong to be always guided by the best: Consequentialism and friendship," *Ethics* 101 (1991).

[2] R.E. Bales, "Act-utilitarianism: account of right-making characteristics or decision-making procedures?" *American Philosophical Quarterly* 8 (1971).

[3] Op cit. note 1, especially Williams, "A Critique of Utilitarianism".

[4] Practical reasons which render their target attitudes *useful* but not *fitting* are often called the "wrong kind of reasons". Cf. Wlodek Rabinowicz & Toni Rønnow-Rasmussen, "The strike of the demon: On fitting pro-attitudes and value," *Ethics* 114 (2004), and R. Y. Chappell, "Fittingness: The Sole Normative Primitive," *Philosophical Quarterly* 62 (2012).

[5] What if an evil demon will punish you for desiring that you have the optimal first-order desire, or for acting so as to bring it about? Then it's desirable (fitting to desire) that you somehow simply acquire the optimal first-order desire (by brute chance, say) without in any way aiming or desiring to do so. Since a desire *for the optimal first-order desire* is itself undesirable (due to the threat of punishment) despite being fitting, this means that it is also fitting to desire *that you not desire that you have the optimal first-order desire*.

[6] See Chappell, op cit. note 4. For a rough intuitive motivation of this idea, note that a connection to warranted agential responses establishes the relevance of normative claims *to us as agents*. What would be the point in

calling something ‘good’, for example, if goodness had no implications for what *we* are warranted in caring about? Free-floating normative claims, that lacked any fittingness implications, would seem not to speak *to us*. But who else is there for normative claims to speak to? It seems they would be left silent, inert, and empty of normative significance.

[7] Compare G.E.M. Anscombe’s worries about the consequentialist’s willingness to consider anything as betraying a “corrupt mind”, in ‘Modern Moral Philosophy,’ *Philosophy* 33 (1958): 17.

[8] Bales, op cit. note 2.

[9] See, e.g., J. L. Mackie, “Rights, Utility, and Universalization,” in *Utility and Rights*, ed. R. G. Frey (Oxford: Basil Blackwell, 1985).

[10] Peter Railton, “Alienation, consequentialism, and the demands of morality.” *Philosophy and Public Affairs* 13 (1984): 142-43. Below, I follow Railton in focusing on egoistic hedonism for simplicity. But the basic distinction between ‘subjective’ and ‘sophisticated’ psychologies should carry over, in obvious fashion, to other (including impartial) forms of consequentialism. Basically, the ‘subjective’ consequentialist agent directly desires the things her theory specifies as good, whereas the ‘sophisticated’ consequentialist agent instead possesses a desire-regulating faculty that will tend to generate in her whatever desires can be expected to best serve to promote the good.

[11] Ibid.

[12] Ibid, 145.

[13] Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1987), 20-21.

[14] One might wonder if Sophu’s resulting pro-suffering desire will be merely instrumental, since it is produced (by her regulating mechanism) as a means to promoting welfare. But we must take care to distinguish conditions on the desire’s *existence* from conditions on its *content* (or, in other words, to distinguish a desire’s persistence conditions from its fulfilment conditions). Sophu’s regulating mechanism ensures that the desire’s existence is

contingent upon its promoting utility, but that doesn't mean that considerations of utility enter into the content of the desire. Sophu is motivated to pursue suffering for its own sake. It's just that her motivations will change when they cease to promote (expected) utility.

[15] At least, this is so on my interpretation of the 'sophisticated' psychology. Elinor Mason, "Do Consequentialists Have One Thought Too Many?" *Ethical Theory and Moral Practice* 2 (1999): 256, suggests an alternative view on which "We can develop new motives from old motives, but only when they are consistent." This would rule out a sophisticated utilitarian acquiring malicious motivations, however beneficial they might be. However, it's not clear to me whether this is meant to be a conceptual constraint on rational agency, or just a contingent empirical hypothesis about how new motivations actually develop in people.

[16] David Gauthier, "Rationality and the Rational Aim," in *Reading Parfit*, ed. Jonathan Dancy (Oxford: Blackwell, 1997).

[17] Derek Parfit, *On What Matters*, vol. 1 (Oxford: Oxford University Press, 2011), 437-39.

[18] This parallels the familiar distinction between 'object-given' and 'state-given' reasons. The dispositional state is a useful state to be in, but the (metaphorical) 'object' of the state—the set of actions it disposes you towards—does not merit such a disposition.

[19] A disposition might improve one's actions overall whilst being predictably detrimental in some specific sub-domain (e.g. a disposition that helps a previously hostile person become more attentive to the emotional wellbeing of those they interact with, but less likely to help the distant needy). In such a case, we presumably do not want to endorse the clearly detrimental (in)actions just because one's *other* actions have improved.

[20] Thanks to Philip Pettit for his assistance in formulating this distinction.

[21] I remain neutral on the question of whether practical reasoning is best understood as concluding in action or intention.

[22] Though it's an important question whether we interpret this as a single monolithic desire for aggregate welfare, or—as I prefer—a plurality of desires, one for *each* sentient being's welfare. See R.Y. Chappell, "Value Receptacles," *Noûs* 49 (2015).

[23] Samuel Scheffler, 'Agent-centred restrictions, rationality, and the virtues,' *Mind* 94 (1985).

[24] Philip Pettit and Geoffrey Brennan, "Restrictive consequentialism," *Australasian Journal of Philosophy* 64 (1986). See also Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011).

[25] Cf. Michael Slote and Philip Pettit, "Satisficing Consequentialism," Proceedings of the Aristotelian Society, Supplementary Volumes 58 (1984). Slote's satisficing consequentialist merely aims to achieve 'good enough' consequences, which makes sense as a practical strategy but seems rather more puzzling as an account of the rationally warranted ultimate aims.

[26] Bernard Williams, "Persons, Character and Morality," 18.

[27] Brad Hooker, *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality* (Oxford: Oxford University Press, 2000), raises the objection in terms of objective rightness. I don't have particularly strong intuitions about objective rightness, as opposed to *what a morally conscientious agent would do*, so I restate the objection here in the latter terms.

[28] G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), claimed that agents will never be in an epistemic position to justifiably break such rules, but we needn't be quite so pessimistic.

[29] In special circumstances, option (ii) may be truly costless, and so there's a possibility that the agent could reasonably undertake such an investigation, and responsibly reach the true conclusion that breaking the rule really is justified in this case. But this won't be typical, and the crucial point for my purposes is just that one's *prima facie* utility judgments won't provide sufficient justification for breaking generally-reliable rules.

[30] This observation may also aid the consequentialist in addressing alleged "counterexamples" to consequentialism, involving organ harvesting, pushing people in front of trolleys, etc. Insofar as we can

accommodate the intuition that there's something wrong with any *person* who would act this way (in any remotely realistic circumstances such behavior would be far too reckless for a morally conscientious agent aware of their own fallibility to rationally perform), the consequentialist may doubt whether we have such a clear grasp of any *further* intuitions about the "objective wrongness" of the irrationally reckless act. (Consequentialists certainly won't feel any inclination to regret the occurrence of the utility-maximizing act, for example.)

[31] Michael Stocker, *Plural and Conflicting Values* (Oxford: Oxford University Press, 1989), 321.

[32] Thanks to Eden Lin, Errol Lord, Sarah McGrath, Tim Mulgan, Martin Peterson, Philip Pettit, Doug Portmore, Christian Seidel, Derek Shiller, Peter Singer, Michael Smith, Helen Yetter-Chappell, the NY *Corridor* reading group, and the University of York's *Mind and Reason* group, for helpful comments and discussion.