Willpower Satisficing

RICHARD YETTER CHAPPELL* ⁱᴰ

University of York

Abstract

Satisficing Consequentialism is often rejected as hopeless. Perhaps its greatest problem is that it risks condoning the gratuitous prevention of goodness above the baseline of what qualifies as "good enough". I propose a radical new *willpower-based* version of the view that avoids this problem, and that better fits with the motivation of avoiding an excessively demanding conception of morality. I further demonstrate how, by drawing on the resources of an independent theory of blameworthiness, we may obtain a principled specification of what counts as "good enough".

1. Satisficing as Constrained Maximizing

Satisficing Consequentialism starts from the idea that, rather than requiring the best, it is enough to do *good enough*. It has been motivated in a few different ways. Economists introduced 'satisficing' as a heuristic to minimize search costs (Simon 1956)—a mere means to maximizing expected value, in effect. Michael Slote (1984) presents it as the radical view that it can be rational to gratuitously pick worse options. But for our purposes, I will understand Satisficing Consequentialism as a view motivated by the thought that it can be permissible to do less than the best, when "the best" would be *too much* to ask of the agent. Thus understood, it is no mere heuristic, but neither does it seek to condone gratuitous optimization failures. Rather, it aims to temper the demands of morality just to the extent required to avoid claims of being "excessively demanding".

The most straightforward form of Satisficing Consequentialism introduces some *utility*[1] *baseline n* that an action must meet in order to qualify as permissible.[2] But, as Bradley (2006) points out, such views are susceptible to permitting the gratuitous prevention of goodness, i.e., where the agent acts so as to bring about *n* value, when they could have instead done nothing and allowed an even greater amount of value

to be produced instead. Such a result is clearly contrary to the motivation for the theory, in addition to being absurd in its own right.

To avoid this problem, we must be more careful about how satisficing enters into the structure of our theory. We don't *merely* want "good enough" outcomes, for those allow for gratuitously preventing even better outcomes. We only want to allow worse outcomes when there's a good reason for it (Turri 2005). This suggests that what we really want to do is to *maximize within constraints*. Satisficing comes in, not at the final step of selecting an option, but rather in setting the (optional) constraints within which we must straightforwardly optimize. There is no set level of utility that counts as a "good enough" outcome. Rather, what's "good enough" is for the agent to do the best they can *either* within the demandingness-moderating constraints, or (if they're willing to go "above and beyond") at whatever greater level of burden they're willing to accept. (This latter clause is necessary to prevent gratuitous wastefulness at a level above what is minimally required.)

We thus find that an adequate satisficing view should adopt the following structure:[3]

> **Constrained Maximizing:** An act A is permissible iff it produces no less utility than any alternative action the agent could perform that is either (i) acceptably burdensome or (ii) no more burdensome than A.

A view of this form must then be fleshed out by specifying what burdens are "acceptable"—in the sense that morality can reasonably require of agents that they endure such burdens—and what is instead supererogatory, above and beyond the call of duty. The next section of this paper will take up this task.

But before we get to that, let me briefly flag two major virtues of Constrained Maximizing over traditional (utility baseline) forms of Satisficing Consequentialism. First, it never condones the gratuitous prevention of goodness. Such action is rendered impermissible by the fact that, *ex hypothesi*, there are better outcomes that you could bring about at no greater burden (perhaps by simply doing nothing). Suboptimal action is only permitted when it relieves the agent of an unacceptable burden, in which case it is clearly no longer "gratuitous". Constrained Maximizing thus yields much more plausible verdicts than other forms of Satisficing Consequentialism.

Second, the view yields a natural account of supererogation as *surpassing the demandingness-moderating constraints to achieve even better results*.[4] Traditional forms of Satisficing Consequentialism are most naturally combined with an understanding of supererogation as simply surpassing the utility baseline, or bringing about an outcome that is better than "good enough". But such a view implausibly implies that (costlessly) lazing back in your chair might qualify as "going above and beyond the call of duty", if a great deal of good can be achieved by your doing nothing at all. This seems absurd. Intuitively, burdenless (in)action may be *required*, if better than the alternatives, but never supererogatory. The Constrained Maximizing approach automatically secures this desired result. If the optimal action involves no burden, and so *a fortiori* is "acceptably burdensome", then it will be outright obligatory. Only unacceptably burdensome actions can qualify as supererogatory.

Such actions are "too much to demand", and it reflects especially well on an agent if they are willing to shoulder such significant burdens in pursuit of morally better outcomes.

The Constrained Maximizing approach should thus prove especially appealing to satisficers and others who wish to moderate the moral demands of consequentialist theory and create room for a plausible account of supererogation.

2. What Burdens?

What are the appropriate "demandingness-moderating constraints" for fleshing out the Constrained Maximizing approach? First we must determine the sort of "burden" against which agents should be protected. The most obvious possibility would be to limit the *welfare* costs imposed on agents. We would then be permitted to choose between the best option that doesn't cost us too much, or the best option at up to whatever greater level of cost we're willing to accept.[5] But I think that understanding demandingness in terms of welfare costs is a mistake. I instead propose that demands on our *willpower* (or mental "effort") are the relevant kinds of burdens. I see three main reasons to prefer this view.

First, compare the following two cases:

> **(Difficult Demand)** You could save many lives in the third world by donating a certain portion of your savings to GiveWell-recommended charities. You've never seriously considered doing anything like this before. Are you required to donate the funds?

> **(Easier Demand)** While sleepwalking, you accidentally set up an automatic payment that will transfer the same portion of your savings to GiveWell-recommended charities. This seeps into your dreams, which helps you to get more used to the idea, though you are surprised to find when you awaken that the scheduled transfer is real. Are you required to refrain from cancelling the transfer?

Accepting either demand would cost the agent the same portion of their savings. Nonetheless, I think that the Difficult Demand is, in the relevant sense, *more demanding*. It requires positive action, and hence action-prompting mental effort, on the part of the agent. It asks them to go out of their way to give up a large portion of their wealth to help others, which they've no antecedent inclination to do, and is contrary to their usual habits and self-conception. That's a very difficult ask. The Easier Demand, by contrast, places no positive demands on the agent: it merely asks them to not intervene while their excess wealth is transferred to save lives overseas. Especially if we stipulate that the agent's dream helped to prepare them for the idea, subjectively "normalizing" the situation, it is easy to imagine them finding it much less psychologically difficult to comply with the demand to refrain from cancelling the transfer. And if so, that would seem to render it a *less demanding* ask.[6]

A second reason to prefer an effort-based account of demandingness is that it accommodates the familiar phenomenon of akrasia in the face of what might be

called "prudential demands". For example: adopting a vegetarian diet, or a robust exercising regime, may be positively beneficial to an agent on net, and yet still be experienced as "burdensome" insofar as large amounts of willpower are required to "go against the grain" and forge new habits of behaviour. But regular exercise is not *costly* to the indolent agent; it's merely *hard*.

The defender of a welfarist account of burdens could respond that there is at least a *proximate* cost here, if the exercise is felt to be painful or unpleasant in the moment, or if the vegetarian misses out on the gustatory pleasure of eating meat. Even if these choices offer long-term net benefits, they could still be costly to the present timeslice of the agent, and perhaps that's enough to explain why the prudential choices seem "demanding". But it's not clear that this response always works, as there are plausibly cases of this form that do not involve even proximate net costs.

For example, a lazy person might be reluctant to exercise even if they know full well that they will feel better—even in the moment—if they do it. We may imagine that the only "unpleasantness" here is in the prospect of bringing themselves to actually do it. Perhaps there can be some inherent unpleasantness in the experience of exerting willpower. But the welfare cost of exerting effort in this way is not plausibly great enough to explain the sense of burden that accompanies prudential demands (especially if this slight cost is soon outweighed by greater benefits). Going to the effort of taking decent care of yourself simply *isn't that bad* for you, in terms of prudential value. But it may nonetheless be difficult, or require significant effort. This strongly suggests that it is the amount of effort required, rather than the prudential cost of the effort, that is the proper measure of the burden being placed on the agent.

It's worth clarifying at this point that the relevant kind of effort here is that involved in exercising the will, not the body. It is the kind of effort involved in getting yourself to the gym, not the physical effort it takes to lift the weights. Physical effort does not seem to have any inherent normative significance, though it may have a kind of instrumental significance insofar as it is experienced positively or negatively (or insofar as it also requires mental effort). But demands on our will—our very *agency*—have, I suggest, the distinctive phenomenology associated with normative demands, in addition to matching up extensionally.

Thirdly, Sobel (2007) has observed that it is difficult to make (non-question-begging) sense of the thought that consequentialism is unusually demanding in terms of the costs imposed by its demands and restrictions. After all, however much it hurts the rich to be required to donate more to the global poor, to allow them to refrain from doing so is surely far more damaging to the poor themselves. So why do we not intuitively hold that non-consequentialist views are "too demanding" on the poor? One natural answer is that the relevant notion of demandingness is more closely tied up with *agency* than with *welfare*. So, while non-consequentialist views may be comparatively more damaging to the interests of the poor, they do not place such stringent demands on anyone's agency or willpower—telling them to *do* difficult things, as opposed to just letting them *experience* difficulties—as maximizing consequentialism does. A willpower-based account of moral burdens thus makes better sense of our ordinary thought involving the concept.

I thus propose that we understand the "demandingness" of moral requirements primarily in terms of their *difficulty*, or the demands placed on our agency, executive control, or willpower[7]—rather than their *costs*, or the demands placed on our wealth or welfare.[8] I will proceed to explore the viability of a view according to which moral demands are understood exclusively in terms of difficulty. (But much of what I go on to say could be adapted to a hybrid theory of burdens that gives both difficulty *and* cost some independent weight. One reason I do not leap to embrace this more complex view is that it raises tricky questions about exactly how to balance the two—a problem that is avoided if, as I believe, the pure willpower account turns out to be adequate.) An appropriate "demandingness-moderating constraint" thus introduces an *effort ceiling X* which we need not surpass in pursuit of better outcomes. The resulting view:[9]

> **Willpower-Satisficing Consequentialism (WSC):** An act *A* is permissible iff it produces no less utility than any alternative action the agent could perform at the cost of up to either (i) *X* willpower, or (ii) the willpower cost of *A*, whichever is greater.

It's important to clarify that WSC doesn't require (or even necessarily permit) you to "try hard" to achieve moral ends. That would be fetishistic. If you can achieve better results with less effort, then you're required to do just that. It merely places a *ceiling* on how much mental effort morality can demand from you. Within that constraint, the requirement is still just to do as much good as possible.

Different theories of this form may be reached by fleshing out the effort ceiling, *X*, in different ways. It plausibly ought to be context-sensitive, e.g. to ensure (1) that it's never permissible to do just a little good when a huge amount of good could be achieved by an only *slightly* more effortful action; (2) that vicious people can't get away with doing little just because it would take a lot more willpower for them to show the slightest concern for others;[10] and perhaps even (3) that your current effort ceiling takes into account your past actions. The next section will show how the best version of WSC appeals to different values of '*X*' for different circumstances.

3. Satisficing and Quality of Will

So far we have seen that Constrained Maximizing—especially in the form of Willpower-Satisficing Consequentialism—is the best form of satisficing consequentialism. But that leaves open the question of whether consequentialists should be satisficers at all. Indeed, I think there is a sense of 'ought'—what we might call "the ought of *most reason*"—of which maximizing act consequentialism gives the right account. But I now want to argue that consequentialists should (additionally)[11] accept WSC as the right account of what's permissible—in the sense of 'permissibility' that is closely tied to blameworthiness, or whether the action is one that a (momentarily fully-informed) competent agent possessing *adequate moral concern* might perform. We will find, along the way, that this exploration gives us the resources to provide a principled specification of the effort ceiling *X*.

I assume a minimal conception of consequentialism as concerned just with the deontic assessment of actions (and perhaps the desires and intentions that underlie them). Other kinds of judgment-sensitive attitudes, such as beliefs and emotions, might be subject to non-consequentialist (e.g. epistemic) norms, and the consequentialist should feel no pressure to deny this (Chappell 2012). As such, it is open to the consequentialist to draw upon the resources of these non-consequentialist norms when appropriate. For example, our account of *expected value* may appeal to independent (non-consequentialist) epistemic norms concerning what credences are warranted in light of the evidence available to the agent. We may thus defer, in part, to epistemological theories. Similarly, I now propose, we may draw upon independent (non-consequentialist) norms concerning when emotions—in particular, reactive attitudes such as blame—are warranted. We may thus defer, in part, to theories of blameworthiness, such as "Quality of Will" accounts that hold feelings of blame to be warranted when the target of the attitude has acted in a way that displays inadequate moral concern (Strawson 1962; Arpaly 2003).[12]

In what follows, I will thus feel free to draw upon the resources of a presupposed Quality of Will account.[13] In particular, I will assume that the Quality of Will theorist can furnish us with a specification of the boundary between adequate and inadequate moral concern, as it appears in their theory. I will then show how consequentialists can co-opt this independently specified distinction to provide a principled specification of WSC's effort ceiling.

But first, back to the general conception of permissibility as what might be chosen by a (momentarily fully-informed) competent agent possessing adequate moral concern.

To flesh out this general concept along distinctively consequentialist lines, we will want to understand *competent agency* and *moral concern* as directed at promoting the welfare of sentient beings, without intrinsic concern for deontological constraints.[14] Even when in possession of all relevant information, including true moral beliefs, as human beings we're apt to fall short of the consequentialist ideal in various ways—most notably, by favouring our own interests (and those of our friends and family), but also by sheer force of habit and general weakness of will (plausibly a major reason why more people aren't vegetarian, for example). Given these facts, it seems natural for consequentialists to understand an agent's degree of moral concern as a combination of (i) the extent to which they value others' welfare, and (ii) their strength of will, or the mental effort they'll put in to actually promoting the general welfare.

That's my proposal for the general shape that moral concern takes, on a consequentialist picture. This provides the first step of our argument: Greater moral concern will go along with a greater tendency to promote the general good. So if we're required to do what's compatible with an *adequate* (but not necessarily maximal) level of moral concern, then this is strongly suggestive of Satisficing Consequentialism: a requirement to do (some) good, but not necessarily the maximum possible.[15]

We can move from Satisficing Consequentialism in general to my willpower-based version in particular by either of two routes. The first simply appeals to our

earlier claim that WSC is the most plausible form of Satisficing Consequentialism, so if we're going to satisfice at all, we should embrace WSC as the best way to do it. The second argument is more interesting, as it makes the case that WSC is especially well-suited to our current conception of permissibility.

Recall that we're understanding moral concern as a combination of altruistic concern and strength of will. So a lack of adequate moral concern may be due to either an excessive degree of self-centeredness or else sheer moral laziness. Either of these might be overcome, in principle, by a sufficient exertion of moral effort. (Of course, the self-centered agent may not wish to put in such efforts. But I assume that it remains within her power to do so—that she *could*, if she wished, override her habitual selfishness and act so as to benefit others instead.) Call that sufficient level of effort—the amount of effort required for this particular agent at this particular time to compensate for their shortcomings and act with adequate moral concern—'*X*'. Doing the best you can with *up to X* effort, as WSC requires, is thus going to be the least burdensome act attaining a level of value compatible with an adequate degree of moral concern, and hence will be the least you can do that qualifies as 'permissible' in the current sense. So this conception of permissibility, together with a consequentialist understanding of moral concern, leads very naturally to WSC.

This also reveals a principled basis for (metaphysically) determining the context-sensitive "effort ceiling" that plays such a crucial role in the formulation of WSC.

> **The Effort Ceiling *X*** (for an agent at a time): that value *X* such that, *in order to qualify as "adequately concerned"* according to our prior theory of blameworthiness and quality of will, the agent must be willing to expend up to *X* willpower in pursuit of the general good.

To illustrate, consider a couple of different ways this might play out. Suppose I start out as a selfish jerk, with minimal concern for others. I'm naturally inclined to act in ways that discount others' interests in ways that it would be reasonable for them to resent. It would take significant moral effort on my part to overcome these bad tendencies and instead treat others in a way that shows adequate concern for their interests, and so the effort ceiling for me at this time is very high. Later, I become a more caring, sympathetic person. I'm still a bit weak-willed, and often tempted to favour myself in ways that would be blameworthy, but it only takes one quarter the willpower that it used to in order to bring myself to act in ways that show adequate concern for others. Accordingly, my new effort ceiling is just one quarter of the previous value. Significantly less moral effort is now required of me. (Though similar minimally-decent *actions* may have been demanded of me throughout.)

As this demonstrates, more moral effort may be required of the less-altruistic, thus resolving the question raised in our previous section of whether vicious people should get away with acting worse simply because it takes them a lot of effort to do the slightest good for others. This account avoids letting the vicious off lightly, because a less naturally-altruistic person is, in that respect, exemplifying less moral concern. So to bring their overall moral concern up to an "adequate" level, a correspondingly greater degree of moral effort is required, to compensate

for their lack of initial concern. We thus find that determining WSC's effort ceiling by reference to prior norms of adequate moral concern allows us to develop WSC in an especially plausible direction.

While WSC thus benefits greatly from being paired with a theory of blameworthiness and quality of will, I don't think the theoretical benefits here are limited to flowing in one direction only. I have so far been assuming that what counts as an "adequate" level of moral concern is a matter to be determined independently, and prior to its incorporation into WSC. But insofar as we accept the methodology of reflective equilibrium, it remains open to us to *also* take our considered judgments about what would be an appropriate effort ceiling for WSC to in turn inform our theory of blameworthiness and (in particular) where to draw the line of "adequate" moral concern. If that's correct, then the further development of WSC may also be of interest to Quality of Will theorists and others working on blameworthiness.

4. Objections and Replies

We may further flesh out WSC by considering how it can respond to various challenges and objections.

4.1 Does WSC demand too much of those for whom self-sacrifice comes easily?
Not necessarily. The naturally altruistic agent will not need to exert much effort at all in order to exemplify adequate moral concern, so their effort ceiling X will be correspondingly low. Granted, if it takes literally *no* effort for them to maximize the good, then they will be required to do so by WSC. But then, for readers on board with the discussion in §2 (suggesting that difficulty is a more relevant metric than cost for assessing moral "demandingness"), it should no longer seem such a burden for a perfectly altruistic agent, as they are merely being asked to act in accordance with their naturally virtuous impulses.

Some may object at this stage that it is nonetheless unreasonable for morality to exploit the selfless agent so, and that the agent really *ought* to give greater weight to their own interests. I do not aim here to address any such sweeping rejection of impartial consequentialist thought. It's important to separate out the strand of the demandingness objection that I *do* seek to address—the strand concerned with the difficulty of acting for the best, and that seeks to accommodate (a reasonable degree of) human weakness—from the *special obligations* objection to impartial consequentialism. For those who think we have special obligations to ourselves or our loved ones, the problem is not that saving five strangers over one friend is "too much" to ask, it's rather (they think) the *wrong thing* to ask. That's not an objection that satisficing can help with (it instead calls for revising our axiology to incorporate agent-relative values), and is beyond the scope of this paper. The sort of "demandingness objection" that motivates this paper instead concerns the distinctive emotional burden of feeling *in conflict* with normatively ideal ends (and the associated demands on our will to resolve the conflict by bringing our intentions into line), rather than fundamental doubts about whether impartial moral theories are aiming at the right ends to begin with.

One might nonetheless hold that there remains another kind of demandingness objection in this vicinity. Suppose you hold that while it'd be morally *ideal* to save five strangers over one friend, still this is "too much" for morality to *demand* of us, even if we could very easily make this decision in the moment. Suppose, for example, that Adam has ingested a pill that temporarily suppresses his ordinary emotions and attachments. Reasoning coldly, he determines that it would be morally better to save the five strangers rather than his best friend, and he feels no immediate emotional resistance to the idea, but he also knows that the drug will soon wear off and he will feel devastated for the rest of his life if he lets his best friend die. Is Adam, while in this coldly robotic state, morally required to save the five, condemning his friend to death and his future self to immense guilt and grief?

I'm inclined to bite the bullet and insist that *if* saving the five really is the morally best act, and it would require no great effort on Adam's part to choose accordingly, then he is required to do it. After all, if it's not the case that we ought to be partial, then the five could reasonably complain that Adam wrongly neglects their interests in acting otherwise. And it's not as though he could plead ordinary human weakness of will in the face of overwhelming emotional attachments. The drug precludes all that. Instead, Adam would be making a cold, deliberate decision to count the strangers' interests for less. And that is (plausibly) objectionable.

I expect that most who reject this verdict will also reject the background assumption of impartialism: that it would be morally ideal to let your friend die in order to save five strangers. If that is not in fact the most morally desirable outcome, then Adam should assign it lower value than the outcome in which he saves his friend. Plugging such an agent-relative account of value into WSC (or maximizing act consequentialism, for that matter) will then yield the desired verdict that Adam ought to save his friend. So it's not clear to me that there is a robust objection to WSC here.

Having said that, if some readers judge this case differently—holding that Adam *ideally ought* to save the five, but nonetheless *needn't* do so—moderate revisions could be made to the account of moral burdens without having to give up on the broader structural views I've defended in this paper. In particular, one could (as noted in §2) adopt a hybrid view of moral burdens as some weighted balance of willpower and welfare costs. But that is not the view I favour.

4.2 Does WSC constrain us to expend our willpower only in maximally morally productive ways?

We earlier noted that prudential requirements can seem burdensome (requiring an effort of will) in much the same way that moral requirements can. It seems that our willpower may also be used in pursuit of various optional projects, such as training for a marathon or practising a musical instrument. Given that such personal projects potentially compete with moral goals that we could be pursuing instead, will it ever be permissible to direct our willpower towards such non-moral ends? After all, whenever we direct our efforts in these ways, there are likely to be higher-utility outcomes that we could have achieved with no greater expenditure

of willpower, thus rendering our action impermissible by WSC's lights. Does this make WSC objectionably "moralistic" and constraining?

This strikes me as the most serious objection to WSC. To address it in detail would be a project more fitting in scope to a book. So for now I will merely sketch what I take to be one promising (but not decisive) line of response.

Consider the hidden costs involved in choice policies governing repeatable situations. While it might seem in any given case *taken in isolation* that we could achieve some important moral goal with no greater effort, this may no longer hold if we imagine trying to implement a lifelong policy of *never* directing our efforts towards non-moral causes. The mental effort and burden involved in forgoing marathon training *as a first step towards forgoing all non-moral efforts* may be very great indeed. And even one who does not explicitly conceive of their decisions in such a way may find the constant moral constraints taking an increasing mental toll over time.

Further, for agents psychologically constituted such that it really is just as easy for them to direct their every effort towards moral goals, we may simply accept the implication that this is what they ought to do. Others' lives really are more important than our hobbies, after all! WSC may rest content with having created some space for ordinary human weakness, and separating the supererogatory from the merely permissible. This much, consequentialists can comfortably achieve. I suspect that further accommodations with "common sense" may come at too great a theoretical (and moral) cost.

4.3 It was claimed that WSC precludes the gratuitous prevention of goodness, but what if an agent perversely wants to prevent good outcomes, and would have to expend effort to rein in this impulse?

This troubling possibility is precluded by our specification of *The Effort Ceiling X* in terms of adequate moral concern. The kind of maleficence involved in wanting to gratuitously prevent benefits to others is clearly incompatible with minimally adequate moral concern. As such, the effort ceiling for the agent will be raised as high as necessary for them to override this maleficent impulse.

I assume that maliciously harming (or preventing benefits to) others is not compatible with minimally adequate moral concern. No decent person could treat harming innocent others—or making the world a worse place—as a worthy end in itself. Suppose it would take E effort for the agent to rein in their malicious desires and refrain from harmfully interfering in this case. Then, by our specification of *The Effort Ceiling X*, X will be at least as great as E—it would take at least this much effort for this agent at this time to act with minimally adequate concern. Plugging this value into WSC, we find that the agent must (at least) perform the best action they can with that costs no more than X willpower. Maliciously interfering will obviously not meet this criterion. They would do better to refrain from interfering, and that would cost just E willpower, which is indeed no more than X.

A complication: Suppose there is some even better action, B, that the agent could perform at this time for less than X effort, but *only if* they redirect their efforts away from resisting temptation and towards achieving this even better result.

(Perhaps they were resisting the temptation to pop a child's awesome bubble when suddenly a charity canvasser comes along asking for last-minute donations towards an ultra-effective project that is otherwise about to close down.) It would then be an implication of WSC that the agent may permissibly give in to their malicious desires in this situation, insofar as they do so as a necessary precondition for achieving *B*. This strikes me as an implication that consequentialists should accept. The resulting bad action is no longer "gratuitous", in context, though of course it reflects badly on the character of the agent that they could not feasibly pursue *B* without acting badly in this other respect.

A different variation on the original objection arises if we imagine that the agent is in no way malicious, but merely has an (intrinsically innocent) impulse to perform an act that will (contingently) have the effect of causing some minor harm to others. It's no longer so obvious that giving in to this impulse is necessarily incompatible with minimally adequate concern for others. If the harm is very minor, and the effort involved in resisting the impulse is very great—such that, for example, the agent would not resent being on the receiving end of the slightly harmful behaviour of someone similarly situated—then perhaps it could be argued that giving in to the bad impulse is no longer entirely "gratuitous" in this context. This could be so if resisting truly would be more trouble than it's worth. But whether such a verdict is warranted will depend on the details of our theory of adequate concern. I'm happy enough to accept this verdict as a (presumably rare) possibility, but it is not a strict commitment of WSC, so others may freely reject it if they find it implausible.

Putting aside these few special (and justified) exceptions, then, we find that by combining WSC with *The Effort Ceiling X*, we may secure the impermissibility of any *genuinely* gratuitous (including malicious) preventions of goodness.

4.4 How does WSC apply to non-consequentialist agents?

WSC's injunction to *do the best you can without having to exceed the effort ceiling* makes no explicit reference to the agent's moral beliefs. But mistaken moral beliefs might be thought to indirectly affect the application of this principle insofar as they affect the effort required to perform various actions.

For example, in Trolley Cases most of us would find it easier to save the five lives by redirecting the trolley on to an innocent person on a side track (Switch), rather than by pushing an innocent person in front of the trolley (Bridge). Depending on how the effort ceiling works out in context, it could plausibly turn out to be obligatory for a typical agent to save the five by killing one in the Switch case, but supererogatory in the Bridge case.

This may seem a surprising result, at least by consequentialist lights. But I think it is the right way to go. Importantly, we are not holding that false moral beliefs themselves legitimate suboptimal behaviour. It is only when so acting would be extremely psychologically difficult for the agent that WSC accommodates failures to so act. It does not generally matter whether one's difficulty stems from false moral beliefs, internalized norms, mental illness, or other human foibles (so long as it does not indicate a problematic quality of will). It is simply the difficulty itself that WSC responds to.

Moreover, once we distinguish moral beliefs from internalized norms, it is more plausibly the latter that typically tend to influence how difficult we find various actions. Even committed Act Utilitarians are likely to be strongly averse to killing in the Bridge case, after all. It is easy to imagine someone saying, "I know it would be best, and I ought to do it, but I just can't bring myself to do it . . ." To overcome such strongly internalized norms in particular cases can be extremely difficult,[16] and so it seems appropriate that WSC mitigates this form of agential burden as it would any other.

4.5 *Doesn't blameworthiness/adequate moral concern depend on a prior account of permissibility, rendering this view circular?*

I'm committed to the view that blameworthiness and adequate moral concern are prior to permissibility. Stated in general terms: there are basic norms concerning when the reactive attitude of blame is warranted, and these norms serve as inputs to determining which acts are permissible. More specifically, WSC's effort ceiling X depends upon how much moral concern is adequate (for an agent in some specific circumstances) to avoid blameworthiness. So while some may seek to reverse this order of explanation and explain blameworthiness in terms of impermissibility, I don't see any reason to think that we *must* do so.[17]

One reason for preferring my order of explanation is that it is less obscure how the binary element is entering the picture, insofar as there is a sharp natural distinction between blaming someone (however slightly) and not blaming them at all (or even positively praising them). So we can understand there being a corresponding normative binary between *warranting* blame or not. By contrast, when considering only actions, it is not entirely obvious where any kind of (non-maximal) binary distinction is supposed to come from—especially for consequentialists (Norcross 2006).

A second reason to prefer explaining (im)permissibility in terms of blameworthiness, rather than vice versa, is that the latter concept is arguably better understood, or easier to grasp independently. *Blameworthiness*, as a concept, is easy to directly grasp. *Permissibility*, by contrast, is more elusive to at least some philosophers—myself included. Whereas Parfit (2011, 165) expresses an "indefinable version of the concept *wrong*" using the phrase "*mustn't-be-done*", others may find no solid grasp on the concept except via connection to some better-understood concept such as blameworthiness.[18] If (im)permissibility turns out to be conceptually reducible to blameworthiness, then there would be no independent (im)permissibility facts eligible to ground the blameworthiness facts.

I don't take either of these reasons to be decisive, and certainly do not expect to persuade everyone to my way of thinking on this point. I merely hope to establish that one may defensibly take blameworthiness to be prior to permissibility, and indeed that many should find this an independently attractive position.

5. Conclusion

In this paper, I've developed three key ideas. First, I argued that, in order to avoid licensing the gratuitous prevention of goodness, Satisficing Consequentialism

should take the form of *Constrained Maximizing*: requiring agents to do (at least) the best they can without undue burden. Second, I argued that the traditional welfaristic conception of moral demands should be replaced by an agential conception which understands moral demands primarily as demands *on our willpower*. Amongst other things, this better accounts for the potential burdensomeness of prudential requirements (e.g., to exercise and floss daily). Finally, bringing these two elements together, I proposed a way to rescue Satisficing Consequentialism: a view that many have assumed to be beyond all hope.

Willpower-Satisficing Consequentialism (WSC) is both extensionally plausible and theoretically well-motivated. Extensionally, it avoids condoning the gratuitous prevention of goodness—a problem that has plagued traditional forms of Satisficing Consequentialism. Theoretically, it meshes seamlessly with a conception of impermissibility that is closely tied to considerations of blameworthiness. Appeal to Quality of Will accounts of blameworthiness provides satisficers with the theoretical resources to "draw the line"—or determine what qualifies as "good enough"—in a uniquely principled, non-arbitrary fashion. What's more: this conception of impermissibility, when conjoined with a distinctively consequentialist understanding of quality of will and "adequate moral concern", yields an independent argument for WSC. We thus find that there is a form of Satisficing Consequentialism that is more than good enough.

Notes

[1] For simplicity, my presentation will assume a welfarist account of the good, but none of the arguments hang on this; non-utilitarians should feel free to make the appropriate verbal substitutions while reading the paper.

[2] For example, n might be some fixed utility value (where, if no available options meet this minimum standard, then the agent is instead required to maximize), or it might be determined in proportion to the best available option in the circumstances—cf. Hurka (1990)'s distinction between *absolute* and *comparative* satisficing.

[3] Rogers (2010) develops a traditionally structured form of satisficing consequentialism, with epicycles added to addresses Bradley's challenge. But I think a more elegant solution would be preferable.

[4] This may be classified as mere 'weak', 'qualified', or 'quasi'-supererogation, as per Horgan and Timmons (2010, 37)'s definition: the acts in question "are at least presumptively morally required, but owing to various mitigating factors, one is justified in failing to perform them (or perhaps excused from performing them), and thus failure to perform them does not call for the sorts of negative reactive attitudes that are appropriate for unjustified or unexcused cases of wrongdoing." I don't think consequentialists should have much sympathy with stronger forms of supererogation, according to which the best act is not even *prima facie* morally required.

[5] Bradley (2006) mentions a view roughly along these lines, attributing it to conversation with Garrett Cullity. Murphy (2000, 64–70) refers to it as a "limited principle of beneficence", but worries that there is no principled way to determine the appropriate limit—a problem I address in section 3.

[6] It's worth clarifying that the doing/allowing distinction is not here assumed to be of any *fundamental* normative significance. As a contingent psychological fact, we often find it easier to *refrain from intervening* than to *go out and do* something. So I exploit these general psychological facts in constructing my scenarios. But if we imagine an agent who faces a situation similar to that of Easier Demand, except that the agent is psychologically constituted such that they find it *just as difficult to comply* as in the Difficult Demand scenario, then we should conclude—both intuitively, and according to the account I'm putting forward—that a requirement to refrain from intervening would be just as demanding for

them as the Difficult Demand would be. It's worth flagging that many may find this an independently appealing account of the (merely derivative) significance of the doing/allowing distinction.

[7] I take this to be the same concept as that discussed under the label of 'ego depletion' by psychologists since Baumeister et al. (1998). But note that nothing in my paper depends upon the truth or replicability of the specific empirical claims they make about ego-depletion.

[8] Eriksson (1997) and McElwee (2016) discuss similar ideas, i.e., that "difficulty" should be distinguished from "cost", and may play at least some part in our understanding of moral demandingness.

[9] I present this as a form of satisficing, but it might alternatively be presented in terms of "agent-centered prerogatives", as per Scheffler (1994).

[10] The next section will spell out one way that this feature might be implemented. To anticipate: the effort ceiling will be raised in proportion to the moral inadequacy of one's quality of will.

[11] It would go beyond the scope of this paper to here defend this form of "deontic pluralism" according to which maximizing and satisficing consequentialisms are not necessarily in conflict. It is a view I plan to defend in future work.

[12] It's important to distinguish the *emotion* of blame from the act of *expressing* blame—the latter, as an act-type, falls under the jurisdiction of minimal consequentialism, but the former, as an internal emotion, does not. More expansive forms of consequentialism might, of course, develop their own— often indirect or 'sophisticated'—accounts of blameworthiness. For present purposes, it suffices to note that one can be a minimal consequentialist, in my sense, without undertaking any such further commitments. I argue elsewhere (Chappell forthcoming) that sophisticated consequentialism is liable to yield incorrect verdicts.

[13] I don't mean to rule out that other theories of blameworthiness might be plugged in here instead— but exploring such possibilities is work for another day.

[14] Deontologists may still demonstrate adequate (albeit slightly confused or imperfect) moral concern, so long as their moral ideology does not prove too costly in terms of welfare (compared to a right-minded but tolerably lazy or self-centered agent). When it does prove excessively costly, however, their moral concern is no longer *adequate*, in the technical sense used here. They may still be considered "well-meaning" in an ordinary sense, and this might even constitute an *excuse*, mitigating the blame that would ordinarily be warranted when acting with inadequate concern. Crucially, however, false moral beliefs do not *justify*—or render permissible—associated actions that would be wrong when performed from non-moral (e.g., selfish) motivations.

[15] Another possibility is to adopt a "dual-ranking" structure, as in Portmore (2011), separately evaluating the *narrowly moral* and *all things considered* reasons (assuming these come apart), and permitting any act that isn't dominated on both scales. But this approach is a non-starter for those, like myself, who see consequentialism precisely as a theory of "all things considered" practical normativity, rather than any more limited domain of "morality", narrowly construed. Without a deep distinction between "moral" and "non-moral" reasons for action, dual-ranking cannot get off the ground.

[16] Especially when they involve salient harms—see Chappell and Yetter-Chappell (2016).

[17] For example, Darwall (2006, ch. 5) similarly explains impermissibility in terms of blameworthiness, and cites a series of like-minded philosophers leading back to Mill (1863).

[18] As McElwee (2010, 314) puts it, "in saying that an agent acts wrongly, we say that (in the absence of extenuating circumstances) he merits blame or serious criticism for his actions."

References

Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford University Press.

Baumeister, Roy F., Bratslavsky, Ellen, Muraven, Mark, and Tice, Dianne M. 1998. "Ego depletion: Is the active self a limited resource?" *Journal of Personality and Social Psychology* 74:1252–1265.

Bradley, Ben. 2006. "Against Satisficing Consequentialism." *Utilitas* 18:97–108.

Chappell, Richard Yetter. 2012. "Fittingness: The Sole Normative Primitive." *Philosophical Quarterly* 62:684–704.

———. forthcoming. "Fittingness Objections to Consequentialism." In Christian Seidel (ed.), *Consequentialism: New Directions, New Problems*? Oxford University Press. (Pending final acceptance of volume by publisher).

Chappell, Richard Yetter and Yetter-Chappell, Helen. 2016. "Virtue and Salience." *Australasian Journal of Philosophy* 94:449–463.

Darwall, Stephen. 2006. *The Second-Person Standpoint*. Cambridge, MA: Harvard University Press.

Eriksson, Björn. 1997. "Utilitarianism for Sinners." *American Philosophical Quarterly* 34:213–228.

Horgan, Terry and Timmons, Mark. 2010. "Untying a Knot From the Inside Out: Reflections on the 'Paradox' of Supererogation." *Social Philosophy and Policy* 29–63.

Hurka, Thomas. 1990. "Two Kinds of Satisficing." *Philosophical Studies* 59:107–111.

McElwee, Brian. 2010. "Should We De-Moralize Ethical Theory?" *Ratio* 23:308–321.

———. 2016. "What is Demandingness?" In M. van Ackeren and M. Kühler (eds.), *The Limits of Moral Obligation*. Routledge.

Mill, John Stuart. 1863. *Utilitarianism*. http://www.utilitarianism.com/mill5.htm. Accessed 13/7/2016.

Murphy, Liam. 2000. *Moral Demands in Nonideal Theory*. Oxford University Press.

Norcross, Alastair. 2006. "Reasons Without Demands: Rethinking Rightness." In James Dreier (ed.), *Contemporary Debates in Moral Theory*. Malden, MA: Wiley-Blackwell.

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

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

Rogers, Jason. 2010. "In Defense of a Version of Satisficing Consequentialism." *Utilitas* 22:198–221.

Scheffler, Samuel. 1994. *The Rejection of Consequentialism*. Oxford: Clarendon.

Simon, Herbert A. 1956. "Rational choice and the structure of the environment." *Psychological Review* 63:129–138.

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

Sobel, David. 2007. "The Impotence of the Demandingness Objection." *Philosophers' Imprint* 7:1–17.

Strawson, Peter. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48:1–25.

Turri, John. 2005. "You Can't Get Away with Murder That Easily: A Response to Timothy Mulgan." *International Journal of Philosophical Studies* 13:489–492.