

# ACCOMMODATING OPTIONS<sup>1</sup>

Seth Lazar

School of Philosophy, Australian National University

## Abstract

Many of us think we have agent-centred options to act suboptimally. Some of these involve favouring our own interests. Others involve sacrificing them. In this paper, I explore three different ways to accommodate agent-centred options in a criterion of objective permissibility. I argue against satisficing and rational pluralism, and in favour of a principle built around sensitivity to personal cost.

## I. Introduction

How should our criterion of objective permissibility accommodate agent-centred moral options? In this paper I consider three possibilities. First, though, I should explain the question and why it matters.

A criterion of objective permissibility is a set of necessary and sufficient conditions for an act's being objectively permissible. 'Objective' permissibility is permissibility in light of all the facts. Not all normative ethical theories offer such a criterion. According to some, the moral landscape is too varied and complex: any such criterion would be an equally complex list of disjuncts and combinations. This is often true of deontological moral theories.<sup>2</sup> Consequentialists, meanwhile, typically do propose a criterion of right action. And they think it an advantage of their view that it can be reduced to a simple set of necessary and sufficient conditions, whereas some popular deontological alternatives

cannot.<sup>3</sup>

This might be a mistake. Deontologists do not *need* a criterion of objective permissibility. Deontologists like Frances Kamm, Victor Tadros, and Jeff McMahan have developed compelling moral theories without ever attempting to cram their views into a simple set of necessary and sufficient conditions.<sup>4</sup> But even deontologists like these might nonetheless benefit from thinking about criteria of objective permissibility, in at least two ways. First, if we can reduce the complex moral landscape to a simpler structural representation, this can help pinpoint some fundamental normative relations, such as distinct dimensions of normative strength. Second, this approach might tame a complex moral theory, enabling us to extend it to decision making under risk and uncertainty.<sup>5</sup>

So, consequentialists already care about having a sound criterion of objective permissibility. Deontologists can defensibly be indifferent, but they also have reason to be interested. What about agent-centred options?

'Agent-centred moral options' here means *options to act suboptimally*: a moral licence to perform an act that is overall morally worse than some permissible alternative. I am interested in two kinds of options: agent-favouring and agent-sacrificing.<sup>6</sup>

I often have an agent-favouring option to prefer my own interests, even when it would be morally preferable to advance the weightier interests of others. At a minimum, I am entitled to favour my own interests rather than advance the slightly greater interests of others. And it is intuitively plausible that I may give my own interests considerable priority.<sup>7</sup>

Whereas agent-favouring options give me licence to prefer my own interests, agent-sacrificing options give me licence to undermine them, even when doing so is overall suboptimal.<sup>8</sup> I am not morally required to act in my own best interest. Many acts that would be wrong if I did them to others, are not morally wrong when I do them to myself. I am permitted both to forgo easily attainable benefits, even, indeed, to harm myself. When I

injure my back at the gym by lifting without proper form, because I'm rushing to get to a presentation, I might act (really) stupidly, but I don't commit a moral wrong.

No doubt there are limits to both kinds of agent-centred options. I cannot favour my own interests by an infinite amount, compared to those of others. And agent-sacrificing options are plausibly limited too.<sup>9</sup> I might not do wrong by inflicting a minor harm on myself, but perhaps it *is* wrong to show total disregard for my own interests, or to abase my interests for the sake of a trivial benefit to others.<sup>10</sup>

Why should we care about accommodating agent-centred options, rather than defending them? Of course we should care about both. In other work, I explicitly defend the view that we have just these kinds of options.<sup>11</sup> But here my task is different. Here we are in the stage of the process of reaching reflective equilibrium where we hold considered judgements fixed, and select between principles to capture their underlying structure. This allows us both to extend our considered judgements about 'easy' cases to harder ones, and to make an informed choice between competing moral theories: opponents of agent-centred options already have well-developed criteria of objective permissibility, as well as first order arguments against those options. Selecting a criterion of objective permissibility for agent-centred options places the competitors on an even footing.

I have a shortlist of three criteria of objective permissibility, all of which were introduced with the aim of accommodating commonsense or conventional verdicts about agent-centred options (among other things). They are sophisticated satisficing; rational pluralism; and my own cost-sensitive criterion. I want to show that the third principle better accommodates agent-centred options than the first two. But I aim to do more than this. A principle that could capture sound pre-theoretical verdicts on cases would be great; one that could illuminate them would be better.

## 2. Satisficing

The basic idea behind satisficing is simple: once you have realised enough value, morality makes no further demands of you. You are at liberty to advance or sacrifice your own interests, provided doing so does not take you below that threshold of value.<sup>12</sup>

Simple satisficing consequentialism has been roundly discredited over the years.<sup>13</sup> And yet, in recent work Jason Rogers has proposed a modified version of the view which appears to answer the standard objections. His main focus was on addressing the criticism—made by Bradley, Mulgan and others—that a satisficing moral theory permits not only gratuitously failing to benefit others, but also gratuitously imposing harms.<sup>14</sup> The objection goes like this: as long as an act is permissible just in case the situation it realises has more value than some threshold, circumstances can arise in which we may actively destroy value, without lowering the situation below the relevant threshold.<sup>15</sup> Rogers proposes the following principle in response:

SAT: There is a number,  $n$ , such that: An act,  $A$ , performed by agent  $S$ , is morally right iff either (i) the value of the situation after  $A$  is at least  $n$ , and is at least as high as the value of the situation prior to  $A$ , and any overall better alternative to  $A$ ,  $A^*$ , is such that: [were  $A^*$  to be enacted instead of  $A$ , either  $S$ 's resultant personal welfare level after the enaction of  $A^*$  would be marginally significantly less than it was prior to the enaction of  $A^*$ , or the value of the situation after the enaction of  $A^*$  would not be appreciably greater than the value of the situation after the enaction of  $A$ ]; or (ii)  $A$  maximizes utility. Rogers (2010: 216)

The first part of (i) precludes the agent from making the situation worse, even if doing so keeps the overall value above the threshold. This deals with the objection that satisficing licenses gratuitous harms. Then the phrase including the square brackets caters for the

worry that satisficing permits gratuitous suboptimality. The agent is required to maximise if doing so doesn't entail severe costs relative to the good achieved, and if the good achieved is 'appreciable'. And then (ii) allows that it's always permissible to maximise utility. According to (ii), if all your alternatives fall below the  $n$  threshold—so none of them would count as permissible on that count—you are required to simply choose the one that maximises utility (the 'lesser evil').<sup>16</sup>

Rogers' goal in developing SAT was to accommodate common sense pretheoretical intuitions about agent-centred options, as well as constraints. No doubt SAT is an improvement on earlier forms of satisficing consequentialism. And yet it still inadequately caters for agent-sacrificing options. It handles agent-favouring options better, but faces an objection on that score too.

SAT rules out actions that lower the value of a situation, even if it remains above  $n$ . So, if I ineptly tear a muscle because I am rushing at the gym, then I have clearly made the situation worse than it was before. What's more, there was an alternative—lifting with proper form—that would have been both better for me and morally better (because it realises more utility). So I have acted morally wrongly. It is standard to think that if I act objectively wrongly, knowing the risk that I would do so, without adequate countervailing reasons, then I am culpable for my wrongdoing. And if I culpably act wrongly, I should feel guilty. So I shouldn't just be angry with myself for having ended my career in the gym; I should feel guilty. This seems seriously out of step with common sense morality. No doubt we can sometimes act wrongly by undermining our own interests—for example, when injuring my back prevents me from fulfilling my other responsibilities. But the mere fact that I have hurt myself does not seem a sound basis for saying that I have acted morally wrongly. Those of us who persistently make irrational choices should not, for the most part, feel guilty as well as stupid. We are not to be viewed in the same light as someone who inflicts the same kinds of setbacks on others. Or at least—that is just the kind of common

sense agent-sacrificing option that it is my goal in this paper to accommodate.

SAT also rules out actions that raise the value of a situation by advancing my interests, when some alternative was available that would have served my interests better. Suppose that I can choose whether to go (on my own) on holiday to Fiji or to Belgium. Going to Fiji will realise appreciably more utility for me than going to Belgium, though going to Belgium would be better than my present situation, and above the threshold. It doesn't involve a cost to my personal welfare—indeed, it makes me better off. So I'm required to go to Fiji. That seems a mistake—again, by the lights of the common-sense agent-sacrificing options that I'm presupposing here are part of commonsense morality. We don't ordinarily think that making irrational choices that affect only ourselves is morally wrong.

SAT also fails to accommodate altruistic self-sacrifice. Suppose that I can choose either to go on holiday to Fiji myself or to pay for your holiday to Belgium. Either option realises enough utility, and more than the status quo ante. But if I go to Fiji, I'll enjoy it appreciably more than you will enjoy going to Belgium. Now suppose that I decide, altruistically, to sacrifice my trip to Fiji so that you can go to Belgium. Is that permissible? Well, no: according to the terms of SAT, there is an alternative act which does not involve marginally significant costs to my personal welfare, and which is appreciably better. So I am again required to maximise. I do something morally wrong by sending you to Belgium instead of taking my own trip to Fiji. Again, this is the wrong judgement. Perhaps if the benefit to you of going to Belgium were utterly trivial, and the benefit to me of going to Fiji would be comparatively enormous, then some will think it morally wrong to sacrifice one's own interests. But this case is different. You'll enjoy Belgium. It's just that I'd enjoy Fiji appreciably more. Sacrificing my own interests here does not seem morally wrong. If I do it knowingly when I could easily have done otherwise, then I shouldn't feel guilt as I wave you off at the airport.

SAT does better with agent-favouring options, but still has one problem: it denies us

agent-favouring options when all our alternatives involve failing to achieve the satisficing threshold. Perhaps, in any choice, there is a value for  $n$  such that at least one act is permissible. But if the threshold is not relativised to the choice situation, as Rogers (2010: 201) implies, then we can face situations in which none of our options is good enough. And in these cases, only (ii) applies. Which means the agent will be required to maximise utility. Which in turn means that she will have to sacrifice her own life, say, if by doing so she can save the life of one person who will be one util better off than her. I think this is no more plausible in these tragic situations, when nothing one does is good enough, than it is when there are satisfactory actions available.<sup>17</sup>

Finally, SAT obviously lacks explanatory power: where does threshold  $n$  come from? How do we determine it? Why is it in one place rather than another? There is nothing in principle wrong with proposing a threshold, but one's theory should offer some way to establish what that threshold should be. Nothing internal to SAT allows us to do so. Perhaps one could adapt SAT to accommodate agent-sacrificing options, and adjust its requirement to maximise when one is below the threshold. However, this would involve introducing further complexity into the principle, and it would not address the concern that the very idea of a threshold is unmotivated.

The cost-sensitive criterion that I propose below has some affinities with SAT, and indeed can pick out a kind of threshold in any choice that counts as 'the least you can do'. But rather than building an unmotivated threshold into the decision rule itself, it gives us a way of working out some sort of baseline. Simplicity is perhaps not the prime virtue of a moral theory, but Occam's razor still applies. If we can convey the same ideas in a simpler way, we must do so.

### 3. Rational Pluralism

The next proposal is more appealing, and has more adherents.<sup>18</sup> I will accordingly devote

more space to it. The basic idea is that, when determining whether an act is morally permissible, we must take into account not only our moral reasons, but also our '(morally relevant) non-moral reasons' for action. Typically these are reasons of prudential self-interest, but they could also include reasons of friendship, conventional reasons, legal reasons, aesthetic reasons, reasons to achieve excellence, and so on. We can account for moral options by arguing that it is sometimes morally permissible to bring about a morally worse outcome, because that outcome is better in some other morally relevant, but non-moral way. The paradigm case, of course, would be when the morally best option goes severely against the agent's self-interest.

I will focus on the most thoroughly developed pluralist proposal: Douglas Portmore's 'Dual Ranking Act Consequentialism'. Here is his proposed criterion of objective permissibility:

DRAC: S's performing x is morally permissible if and only if, and because, there is no available act alternative that would produce an outcome that S has both more moral reason, and more reason, all things considered, to want to obtain than to want x's outcome to obtain. (Portmore (2011: 4))<sup>19</sup>

DRAC is carefully formulated to avoid a number of objections and controversies that do not concern me here. For my purposes, we can gloss this principle as saying the following: an act is morally permissible if and only if no alternative act is both morally better and all things considered better. On this account, agent-favouring options are instances in which the agent-favouring is morally outranked, but all of the morally better options are all things considered worse than, or at least on par with, the agent-centred option, because of the cost to the agent. Agent-sacrificing options, by contrast, are instances in which the agent-sacrificing option is all things considered outranked by other options, but those options are not morally better, so the agent-sacrificing option is morally permissible.

DRAC is the product of an exhaustive engagement with the philosophical literature on consequentialism and commonsense morality. It has many virtues, and it can indeed often accommodate agent-sacrificing and agent-favouring options. Often, but not always. In particular, DRAC cannot adequately accommodate agent-sacrificing options when there is independent moral reason to advance one's self-interest. And it forces us to think of moral options in terms of opportunity rather than production costs, when that is a substantive question that should be left open to debate, rather than foreclosed. I will develop each point in turn.

Any approach to options with this form has to distinguish between moral and non-moral reasons. This is no easy task. Portmore argues that a moral reason is a fact that, morally speaking, counts for or against some action. He adds that a moral reason 'is a reason that, if sufficiently weighty, could make an act either obligatory or supererogatory'.<sup>20</sup> And he identifies one class of reasons that are *not* moral: the agent's self-interest is not, *for that agent*, a moral reason (it is, of course, a moral reason for others). He does not rule out that one might have some *other* moral reason to promote one's self-interest. And he emphasises that sometimes non-moral reasons of self-interest can determine which acts are morally permissible (because, for example, the morally preferable act has excessive personal cost). But he does insist that the mere fact that some act advances one's self-interest cannot count in its favour, morally speaking.<sup>21</sup>

This move is crucial. Without it, the appeal to all things considered rationality would not preserve important judgements about agent-sacrificing options, because the non-sacrificial alternatives would be both morally and all things considered better.<sup>22</sup> Portmore would agree with my verdict on the Fiji/Belgium case above. But if reasons of self-interest were moral reasons, then he could not do so: going to Fiji would be both morally and (because of my reasons of self-interest) all things considered better than paying for you to

go to Belgium. If reasons of self-interest are not moral reasons, then I am permitted to pay for you to go to Belgium, even though my going to Fiji would be all things considered better, because it would not be morally better than sending you to Belgium.

The success of the pluralist move, then, depends on its insistence that merely favouring the agent's self-interest cannot make an outcome morally better. And this is its greatest weakness. My interests are not the only thing about me that matters morally—but they do matter morally, even if I am the one acting. The contrary view deprives us of a whole species of justification: according to Portmore, one act simply cannot morally outrank another *in virtue of* contributing to the agent's self-interest. So if I choose an option that favours my interests, it is not (at least not for that reason) morally better than the self-sacrificing alternatives. It might still be permissible. But it is morally worse.

I think this is a mistake. Sometimes I have a moral justification for acting in my own interests, not merely a rational one. Of course, Portmore agrees that this can be true: sometimes one's other moral reasons tell in favour of advancing one's self-interest. But this combination of views generates a dilemma. Consider these examples:

Suppose we find some manna from heaven, which will bring me 100 units of happiness if I take it, and you 10 units if I leave it for you to take. It seems morally permissible for me to take the manna (I will support this claim in a moment). Portmore can reach this conclusion in one of two ways. First, he can argue that it would in fact be morally preferable for me to let you have the manna, but taking it for myself is nonetheless permissible, because the personal cost to me of doing the morally better thing is so great that the morally suboptimal choice is in fact all things considered better. By taking the manna, I am acting on a kind of right to be selfish. This is not an appealing interpretation of the case. Assuming equal starting points, if an indivisible good to which nobody has a claim appears, it's morally best that it go to the person whom it would benefit the most. This principle—do the most good you can when nobody has a prior claim—seems like a

basic element of commonsense morality, as long as we assume equal starting points.

Now, Portmore *could* respond by agreeing that there is independent moral reason to maximise the good realised by manna from heaven, and in this case that reason happens to coincide with my self-interest. One wonders what that independent reason would be, as distinct from the reasons given by the advancement of my good; but set that aside, because a further problem awaits. If taking the manna is indeed morally best, and if it is also in my interest, then I am morally required to take the manna. Giving it up would be impermissible, because that option, both morally and all things considered, is outranked by the alternative. Taking the manna is better for me, and better morally, than giving it up. But this also seems obviously mistaken. Even if the manna would create more good if I took it, it's still *my* good, so if I'd rather forgo it, I should be able to. Again, no doubt there may be limits to our licence to sacrifice our own interests. But they don't kick in for cases like this. If I simply want to be generous, and let you catch a break, then I'm not doing anything wrong.

We can multiply cases with the same structure. Suppose, for example, that instead of manna we must distribute resources that I produced through my own labour and ingenuity. I take it that commonsense morality would say that I have a prior claim on these resources, other things equal, so if I choose to keep them, my action is sanctioned by positive moral reasons—it is not merely a licensed act of selfishness. And yet if I want to be generous, and give them to you, I am permitted to do so. DRAC cannot accommodate this conventional pair of positions. It must either say that it's morally best to give you the resources, but permissible to keep them because of the personal cost of not doing so (licensed selfishness). Or else it's morally best for me to have the resources, and so impermissible for me to give them to you, because that option is both morally and all things considered outranked.

Or now suppose that I can choose between two distributions, A and B, of some manna

from heaven, to which nobody has a prior claim. Again, all potential recipients are at equal starting points. A divides equally between all, while B gives everyone else a little more and me the corresponding amount less. Again, Portmore could argue that B is morally best, but A is nonetheless permissible because it outranks B all things considered. But this too seems wrong. If everything else is equal, then the egalitarian distribution *is* morally best. This is not, I think, controversial. Even those who are hostile to egalitarianism in fully-developed theories of distributive justice tend to agree that we have *some* reason to realise egalitarian distributions, at least when doing so works to *someone's* benefit.<sup>23</sup> Portmore I think would agree on this point. So he might then argue that A is indeed morally best—but this would make it morally required. B is outranked by A all things considered, and now morally. So I am not entitled to sacrifice my own interests to advance those of others, because doing so would undermine the equal distribution.

Finally, consider a case of self-defence. Suppose a culpable attacker threatens the life of an innocent defender. A popular justification for killing in self-defence goes like this: someone has to die, and the defender must choose between killing and being killed.<sup>24</sup> There is a strong presumption against killing another person to save oneself, which has different grounds in different theories, but is widespread within conventional morality and all legal systems. No legal system in the world would view the fact that by killing another I saved my own life as a justification that would exclude trial for murder.<sup>25</sup> To override that presumption, we need a moral asymmetry between the two people whose lives are at stake. In this case, if the defender saves herself, she saves an innocent person; if she kills the attacker, she kills someone who is not only culpable, but culpable for this very situation arising. The fact that the defender can save her own life is therefore crucial to justifying killing the attacker. But if the defender's survival cannot make the outcome morally better, because she does not have moral reasons to act in her own interests, then what can morally justify lethal defensive force?

Again, Portmore might respond that, even if killing the attacker is not morally justified, it is all things considered rational, and therefore morally permissible. But even if this gets the right deontic verdict, it does so for the wrong reasons. Killing a culpable attacker in self-defence is *not* a morally suboptimal outcome that one is entitled to bring about because the cost to you of not doing so is too great. The defender acts justly, *and* brings about a better outcome, when she saves her own life. The culpable attacker's interests are discounted by his culpability; the defender is innocent, so her interests are not discounted.

Again, one could argue that the defender has reasons of justice to kill the culpable attacker, but that too would imply that she is required to do so, which is very likely false. If she wants to let herself be killed, then—at least assuming she has no outstanding obligations to others—she may.<sup>26</sup> What's more, it also implies that there is some positive reason for her to kill the attacker, such as retribution for his wrongdoing. And most people think that retribution (and indeed desert, to which it responds) have no place in a plausible theory of self-defence.<sup>27</sup> The point is not that 'it's a good thing' to kill the culpable attacker. It's simply that *it is better* to kill him than let herself be killed.

The same problem applies when considering the infliction of proportionate harm on the innocent as a side-effect of saving oneself from some dangerous threat.<sup>28</sup> Suppose that the defender can save herself only by tossing a grenade, which is likely to injure but not kill an innocent person standing near the culpable attacker. If the defender is justified in doing so, it must be because the outcome in which her life is saved is morally better than the one in which she is killed and the bystander avoids that additional harm. To justify her action in the right way, we have to include the defender's interests in the calculation of proportionality. But if it's both better for her and morally better to toss the grenade, then the defender is required to do so. But it's not plausible that we're required to save our own lives in cases like this, even at the cost of injuring an innocent bystander.

DRAC can account for agent-sacrificing options in many cases. But it runs into

problems when one's self-interest is aligned with one's other moral reasons. One faces a dilemma: either one argues, implausibly, that acting in one's self-interest is like acting on a right to do wrong—a licence to be selfish; or else one claims that one is required to act in one's own interests. Nor is this problem unique to DRAC—when rational pluralism is paired with a consequentialist moral theory, the same problem will arise. If prudential reasons are moral reasons, then we lack agent-sacrificing options to act suboptimally—we're required to act on our reasons of self-interest. If prudential reasons are not moral reasons, then outcomes that seem clearly morally best (and whose moral ranking is an important part of their justification to others) must instead be represented as morally worse than the self-sacrificing alternative, permissible only because the agent has a licence to be selfish.

As with agent-sacrificing options, DRAC can accommodate most agent-favouring options. Its shortcoming is to rule out, without argument, an understanding of agent-favouring options that is worth independent investigation and discussion.

When thinking about agent-favouring options, we have to compare the good one can realise with the personal cost of bringing about that good. But we can understand that cost (and that good) in different ways. On one approach, we consider each option in isolation from the others, and compare it with what would happen if one did nothing. For example, suppose that you could save five lives by entering a burning building, suffering moderate burns in the process. On this approach, we can work out whether you have an option not to enter the building by simply comparing this option with the alternative of doing nothing. The cost to you is the moderate burns; the benefit realised is the five lives saved. Let's assume that you're not morally required to suffer moderate burns in order to save five lives. I'll call this the production costs approach, using a metaphor from economics. Production

costs are simply the personal costs of producing the good.

But there's another way to think about agent-centred options. We could look at opportunity costs instead of production costs. Again, the metaphor is from economics. If we take this approach, then instead of looking at the absolute costs and benefits of an option—which we determine by comparison with a counterfactual baseline in which you do nothing—we instead look at their comparative costs and benefits. If your only options are to do nothing or to save the five, then the opportunity and production cost approaches give the same verdict. But suppose now that you have a third alternative. You could enter the building, save those five, and save a further five lives, at very slight additional cost to yourself—you'd suffer some light bruising as well as the moderate burns. The opportunity costs approach forces us to compare each alternative with all of the others. And it might well lead to the conclusion that (1) it's permissible not to go in, since one is not required to suffer moderate burns and light bruising in order to save ten lives, and (2) it's permissible to go in and save all ten, but (3) it's impermissible to enter the building and save only five.<sup>29</sup>

In the rest of this paper, I've focused on whether DRAC and SAT can adequately accommodate intuitive verdicts that seem to be conventional or common-sense. In this case, it's a little hard to say which of the opportunity or production costs approaches captures the conventional wisdom. Theron Pummer thinks that the conventional wisdom likely supports the production costs approach, and I'm inclined to agree.<sup>30</sup> On this view, if an option is supererogatory, then it's genuinely morally optional. An act that would be supererogatory if you had no other alternatives besides doing nothing cannot become impermissible because a further, better alternative becomes available.

Now, I agree only that this is the conventional wisdom. I think the truth of the matter is more complex, and I don't need to take a side in this debate here.<sup>31</sup> I want only to insist that it's a mistake for our criterion of objective permissibility to foreclose this substantive moral question. And DRAC makes that mistake: it can accommodate only the opportunity-costs

version of agent-favouring options. It forces us to compare act alternatives, and look only at the marginal differences between them. It cannot accommodate the idea that some act might be optional not because of how it compares with all the available alternatives, but simply because the good realised comes at too high a price to the agent. DRAC insists that if you save one person from a burning building when you could have saved two at the same degree of risk to yourself, you have acted impermissibly, even if your action would have been heroic had that alternative been unavailable. Perhaps this, ultimately, is the right way to think about agent-favouring options. But it is not *obviously* right. The alternative is not implausible. The fact that DRAC lacks the flexibility to accommodate the 'production costs' view is a cause for concern.

#### 4. Cost-Sensitivity

Agent-centred options are tricky. It is very hard to incorporate them into a simple decision rule. I suspect that it is impossible to adequately represent them in any satisficing or maximising framework, even those that, like Portmore's, are multidimensional. Instead, I think we need to explicitly build options into the decision rule itself. Here is my proposal:

COST: An act is permissible if and only if either (a) there is no morally better act that has reasonable marginal costs to the agent or (b) it falls short of every such reasonable alternative only in virtue of costs borne by the agent.<sup>32</sup>

COST has many moving parts, all of which need to be explained.

First, it is neutral on what counts as an act. It should be equally acceptable on different interpretations—for example, one might think of acts atomistically, as the smallest elements of agency over which one has voluntary control; or one might think of them holistically, as compounds of those smaller elements, indeed as whole plans or sequences.<sup>33</sup>

COST offers necessary and sufficient conditions for an act being objectively permissible, but it does not insist that fulfilling these conditions is what explains an act's

permissibility (unlike DRAC). This is important. We can view COST as representing another moral theory or we can view it as identifying the grounds of morally permissible action. I take no standpoint on that question. My own interest in COST is instrumental: I think it can help represent deontological moral theories in a manner that renders them amenable to extension to decision-making under risk. I *don't* think that what explains an act's objective permissibility is that it satisfies COST.<sup>34</sup>

This also explains why COST says 'no morally better act'. First, what makes an act morally better is left open: where DRAC insists that outcomes are what matter, COST is neutral (of course, DRAC could be reframed to be more neutral also). Perhaps outcomes are just one factor that influences whether one act is better than another. Nor do I have much to say on what 'moral' means here. One act is morally better than another just in case one has more moral reason to perform it. Our moral reasons can, of course, be diverse. In particular, we can recognise both agent-relative and agent-neutral reasons for action as bearing on our decision. Agent-neutral reasons have the same force regardless of who is acting. The force (and indeed existence) of agent-relative reasons depends on something specific about the person acting.<sup>35</sup> One act might be better than another despite being worse with respect to agent-neutral reasons, because it is better with respect to agent-relative reasons.

What makes these reasons moral ones? That question deserves a paper—a book—on its own, and cannot be quickly answered. But I can sketch how my view would go.<sup>36</sup> Moral reasons are reasons such that, if one acts on or contravenes them, one can be an appropriate object of praise or blame respectively. Of course, other conditions might have to be met—for example, perhaps your action must be all things considered impermissible for you to be blameworthy; and perhaps you need to know that you are acting on or contravening the reason in order to be praise- or blameworthy; and perhaps some circumstances, such as duress, can render even knowing breach of a moral reason

blameless. It's also worth remembering that not all moral reasons attract *both* praise and blame. For some moral reasons, you can be blameworthy for contravening them but not praiseworthy for acting on them (e.g. your reason not to kill an innocent person for your own benefit). For others, you can be praiseworthy for acting on them but not blameworthy for contravening them (e.g.—perhaps—some reasons of courtesy).<sup>37</sup> The point is simply that we can identify moral reasons in particular by their propensity to attract judgements of praise and blame, in the right circumstances. This—especially blame—clearly differentiates between moral and prudential reasons.

The rubber really hits the road in the remainder of clause (a): 'has reasonable marginal costs to the agent'. Each element of this needs unpacking. It will help to start with costs. These can be understood more or less expansively. They might be setbacks to anything that the agent values. Or they might be setbacks to the agent's self-regarding interests.

The first approach appeals because, typically, we think that our agent-favouring options give us a licence not only to advance our own interests, but also to advance the interests of those we care about, and our personal projects more generally.

The second approach appeals insofar as we think that we can accommodate those other kinds of options by focusing on the cost *to the agent* of not being able to help her children, pursue her projects, and so on, as well as by recognising our agent-relative reasons to contribute to those personal projects and commitments. We might also think that agent-favouring options are ultimately grounded in the idea that individuals are ends in themselves, and that this is inconsistent with their being morally required to sacrifice their interests just in case doing so realises more good overall.<sup>38</sup> This would imply that options are grounded in the agent's special authority over her own interests, which she plausibly has only over her self-regarding interests.

However, COST does not foreclose this question. It can operate with a more expansive understanding of costs in clause (a), though it will need a narrower one in (b). On the

narrower interpretation, COST can help explain an interesting distinction in the kinds of reasons we might have for acting partially towards those we care about. Suppose, for example, that a trolley is headed towards five strangers, and I can save them only by diverting it down a side track, where it will kill my son. I think that it would be morally wrong to divert the trolley, given my agent-relative reasons to protect my son, grounded in the value of our relationship.<sup>39</sup> Not diverting the trolley is morally best. But now suppose that there are 50 or even 100 people at risk of being killed by the trolley. Then it might indeed be morally best to turn the trolley. If the person on the side track were a stranger to me, I would be required to pull the lever. And yet, since it's my son, the cost of pulling the lever for this end is surely an unreasonable one for me to have to bear. Even if it would be permissible to turn the trolley in this case, it could not be required. And if the number of people set to be killed by the trolley is high enough, then surely I *can* be morally required to turn the trolley towards my son. Of course, I would not do so, and we would probably think that my action is morally wrong, but significantly excused on account of the severe duress under which I find myself.<sup>40</sup>

The point here is not to defend these specific intuitions about these cases, still less the specific numbers. Your intuitions may differ. The point is instead to mark out the different ways in which COST allows us to understand partiality. Sometimes when we act partially, we plausibly do the best thing. Sometimes we act permissibly by being partial, though acting impartially would be morally better. And sometimes we act impermissibly by being partial, though perhaps our exigent circumstances might mean that we can hardly be blamed for doing so. And of course there is a further possibility, in which we act impermissibly by acting partially, and *can* be blamed for doing so. If I were to kill a stranger in order to secure his kidney for a life-saving transplant for my child, then I would surely be acting impermissibly, and without excuse. One virtue of COST is to bring these different possibilities to the forefront.

In working out whether an act,  $\phi$ , is permissible, we need to compare it with the alternative acts. If  $\phi$  is morally better than all alternatives, then it is permissible. If some alternative,  $\psi$ , is morally better than  $\phi$ , then  $\phi$  is still permissible, provided  $\psi$  involves additional costs to the agent that are unreasonable in relation to the moral gain. So we need to know more about what 'reasonable' means here, and also about 'marginal'.

COST brings to the foreground at least two key dimensions of normative strength, which our other principles either overlooked or obscured. The first is that of overall moral betterness, the second that of 'reasonableness'. These correspond, in Frances Kamm's terminology, to the 'precedence standard' and the 'efforts standard'.<sup>41</sup> In terminology that I have used elsewhere, the former concerns the *gravity* of our moral reasons, the latter concerns their *stringency*.<sup>42</sup> And in Joshua Gert's terminology, they plausibly correspond to the justifying and requiring dimensions of normative strength respectively.<sup>43</sup> One dimension concerns what makes one act morally better than another, the other whether an act is morally required. The key insight captured by COST is that these two dimensions are distinct from one another: one cannot infer the stringency of one's reasons to  $\phi$  from its moral betterness:  $\phi$  might be worse than  $\psi$ , and yet one might have more stringent reasons to  $\phi$  than to  $\psi$ . It is a virtue of COST that it points up this contrast.

COST is framed in terms of opportunity costs. But we could equally well develop a version invoking production costs instead:

COST\*: An act is permissible if and only if either (a) there is no morally better act that has reasonable costs to the agent or (b) it falls short of every such reasonable alternative only in virtue of costs borne by the agent.

This provides the flexibility to accommodate the two distinct understandings of agent-centred options, without foreclosing that discussion.<sup>44</sup> I think opportunity costs are already

well-defined. But there are interesting questions to answer about production costs. One could measure them by comparing the agent's level of well-being before she  $\phi$ s with her well-being after she  $\phi$ s. But while this might sometimes be a useful heuristic, it cannot always be the right way to proceed. Suppose that, whether she  $\phi$ s or not, she will suffer some significant loss. Then we can hardly count as part of the cost of  $\phi$ ing the total cost relative to her antecedent level of well-being. We must instead take an appropriate counterfactual baseline for measurement, accounting for what would have happened to her had she done nothing. I reserve further discussion of this point for a separate occasion; here all I need note is that COST remains open on just how we should calculate production costs.

COST and COST\* can adequately accommodate all agent-favouring options. They are, of course, highly schematic principles. But they do illuminate those options to some extent, by drawing attention to the two dimensions of normative strength, and making clear where the crucial decision points are in one's theory of agent-centred options. Should we understand costs narrowly in terms of the agent's self-regarding interests, or broadly in terms of all those things she cares about? Should we focus on opportunity costs and benefits, or production costs and benefits?

Additionally, COST can illuminate the idea that there is a threshold of moral worth: acts above that threshold that are not gratuitously suboptimal are permissible. In other words, it can provide the threshold that the satisficing approach posits, without motivation.

Clause (a) entails that, in any decision problem, there is at least one option that constitutes 'the least you can do'. This is the morally best act that has reasonable marginal costs relative to all morally worse alternatives. That is, if there are morally worse alternatives that are better for you, you *can* be required to bear that additional cost in order to realise the additional moral benefit. Conversely, any morally better alternatives must

involve excessive personal costs, so that you can't be required to bear that additional cost for the sake of the additional moral benefit.

An example might help here. Suppose that you're outside a burning building. There are three entrances to the building. If you stay outside, and do nothing, then you'll be fine, but ten strangers will die. If you enter door A, you'll save one person's life, at no cost to yourself. If you enter door B, you'll save that person and three others, at some non-trivial but not too serious personal cost—moderate bruising to your upper body, say. And if you enter door C, you'll save all four of those, as well as six other people, at significant personal cost—full-body third-degree burns.

Let's use the opportunity costs version of COST (that is, not COST\*). We can work out the least you can do, by figuring out the morally best option that has reasonable marginal costs compared to every morally worse alternative. Given the way I've described the case, that option is taking door B. Doing nothing and entering door A are both morally worse, because more lives are lost, and the difference in cost to the agent is relatively slight. And the marginal costs to the agent are reasonable given the additional good that can be done. This is true when we compare entering room B with doing nothing. But it's also true when that's compared with entering door A. You save the same person, as well as three others, at only a moderate personal cost—one that, by hypothesis, you can be required to bear in order to save three additional lives. But while entering door C is morally better than entering door B, it is not better by enough to make the significant additional personal cost morally required. Again, it's not important that you agree with my specific intuitions about these cases—what matters is the structural relationship between them, and the way in which COST can tell us what the least you can do is—as well as what counts as going beyond the call of duty. SAT needs to import an arbitrary threshold into this case, to determine whether entering door A or B counts as good enough.

Of course, like SAT and DRAC, COST is highly abstract, and needs to be fleshed out

with substantive moral theory. But that is true for all of them. With COST we at least know how to establish the least you can do—find the best option that has reasonable costs compared to all the morally worse alternatives.

On, then, to clause (b). Clause (a) entails that we may sometimes sacrifice our interests for the sake of the greater good. Clause (b) ensures that we are permitted to sacrifice our interests, even when doing so is suboptimal. The second part of the clause is the easiest to unpack: (b) is plausible only if the costs at stake are narrowly interpreted in terms of the agent's self-regarding interests. One does not, for example, have a licence to thwart the interests of one's children just because one has special reason to care for them. This means that if clause (a) is interpreted using an expansive conception of 'costs', we will need to explicitly flag that costs are understood differently in the two clauses.

The first part of clause (b) is trickier. For a self-sacrificing act to be permissible, there must be no alternative act for which all of the following is true: it has reasonable marginal costs to the agent; it is morally better; and its moral betterness does not reduce to a benefit to the agent. If there is even one act for which these three conditions are true, then the act of self-sacrifice is impermissible.

But there is room, here, for interpretation. COST states that an act is permissible if 'it falls short of every reasonable alternative only in virtue of costs born by the agent'. But what does 'in virtue of' mean here? There are two possibilities, and the extent to which COST improves on DRAC depends on which we endorse.

One possibility is this. Consider a candidate self-sacrificing act  $\phi$ . If an alternative to  $\phi$  is both better for the agent, and morally better in some respect, then  $\phi$  is impermissible. But this is just what DRAC said, which led to a number of counterintuitive implications. Below I introduce a preferable interpretation of COST. But even when understood in this way,

COST remains in better shape than DRAC. It can still handle some of the cases above, where the only thing that explains why one act is morally better than another is that it serves the agent's interests. This is true, for example, in the original manna from heaven case, as well as in the self-defence case.

In the first, if I take the manna I will get 100 units of well-being; if I leave it for you, you will get 10. COST can recognise that it is morally better that I take the manna, and so permissible for me to do so, just in virtue of the fact that my interests matter morally. But this also means that if I want to sacrifice my interests, then I am entitled to do so. So the difference between taking the manna and leaving it comes down only to the difference in my well-being. So it is permissible for me to leave the manna, under clause (b). Because DRAC cannot recognise that my interests can give moral reasons in their own right (since doing so would mean one would be morally required to serve one's own interests when doing so was overall optimal), it cannot accommodate this judgement.

The same is true in the self-defence case. I am permitted to kill a culpable attacker in self-defence because the goal of saving my innocent life is morally worthwhile. But, ex hypothesi, the moral difference between my killing him in self-defence and allowing myself to be killed comes down to the costs that befall me: the loss of my life. And if I want to bear that cost, then I am entitled to do so.

Some of the other cases, however, are trickier. Recall the distributive justice case, in which my self-sacrifice undermines an equal distribution. In this case, my refusing the equalising benefit is worse not only in respect of the cost to me, but also in respect of the damage done to the egalitarian distribution. So COST would seem to agree with DRAC that it is impermissible to sacrifice my interests.

I think that we can solve this problem by interpretation COST differently—and that the process of doing so affords genuine insight into our agent-sacrificing options. We should, I think, adopt a different interpretation of 'in virtue of'. If the difference between one act and

another *supervenes on* the well-being of the agent, then plausibly the agent is licensed to choose the act that is worse for her. This covers cases in which the only difference between the two acts has to do with the agent's well-being; but it also covers those in which there is a further morally relevant difference, but that further difference supervenes on the difference in well-being.

Consider the distributive justice case: the only material difference between the two outcomes has to do with the agent's well-being. The shift from an egalitarian to an inegalitarian distribution supervenes on that difference. So even though the self-sacrificing act falls short of the egalitarian act both with respect to the agent's well-being and with respect to the inegalitarian distribution, it falls short only in virtue of the shortfall in the agent's well-being, since the change in the distribution supervenes on that shortfall.

The basic underlying idea is this: if the moral difference between two acts depends only on *my* well-being, then it should be up to me (perhaps within limits) which of those acts to choose. If the egalitarian distribution is realised only by my well-being being affected, then it should be within my power to determine whether that distribution is realised. The moral difference is made through an operation on something that is mine. So it's up to me whether to realise that moral difference.

COST might now face a different objection. Perhaps the verdicts we are left with would be too extreme: many people think that we have self-regarding duties, and it might seem that this interpretation of clause (b) rules that out.<sup>45</sup> I think there is a problem here, but it is not all that serious. Clause (b) rules out only that one could have a self-regarding duty to promote one's own well-being. Provided the self-regarding duty supervenes on something other than the agent's reduced level of well-being, it can still generate a moral reason that blocks self-sacrifice. For example, (b) is consistent with the view that entering into certain kinds of relationships can be morally wrong, even if all the costs fall on the person who chooses to be exploited. In such a case, the difference between one act and the other would

not simply come down to the reduction in one's well-being, but would also lie in the creation of a morally objectionable exploitative relationship. The same can be said for many kinds of self-abasement. Subordinating one's needs to the needs of another can be wrong because of the kind of relationship that it generates, not simply because of the reduction in the worse-off party's well-being. Or it can be objectionable because it involves taking an inappropriate attitude towards oneself, viewing one's own needs as being less important than those of others.

And yet, still, some people would think that there are limits to one's licence to thwart one's own interests. Clause (b) could be modified to accommodate this view; we would simply need to build into it something reflecting those limits.<sup>46</sup>

What about cases in which one act is worse than another in virtue of both my well-being and something that my well-being is causally responsible for? Suppose, for example, that Alice is a much-loved actor, whose death will make millions of people very sad. We might reasonably think that she is still morally permitted to take her own life, because those millions have no right that Alice stay alive just to avoid making them sad. However, perhaps cases like these can be catered for by clause (a). Alice's death is *not* outranked by another option that has reasonable marginal costs, because she cannot reasonably be expected to forgo exercising her free choice in such a fundamental matter as whether she lives or dies, just to ensure that her fans are kept happy.

## 5. Conclusion

I have argued only that COST better accommodates agent-centred options than SAT and DRAC. I have not shown that there is no possible satisficing or rational pluralist principle that could do the same. Nor have I explored the other strengths and shortcomings of COST. I have sought only to introduce and defend a simple, illuminating principle that successfully accommodates both agent-favouring and agent-sacrificing options—or does

so, at least, better than two prominent alternatives. Of course, any attempt to present a simple and succinct set of necessary and sufficient conditions for an act's being objectively permissible will, unavoidably, be very abstract. This is true for all three principles considered in this paper. Of these three principles, only DRAC has been worked out in great detail, with its abstract bones furnishing the structure for a detailed moral theory. And yet at these high levels of abstraction we can see important structural relationships between moral concepts. My contention in this essay is that COST better captures the underlying structure of commonsense morality than do these two competitors. The task of filling it out into a comprehensive moral theory lies ahead.

### *Bibliography*

- Barry, C. and S. Lazar (2018) 'Moral Options and Optimisation', *Unpublished MS*.
- Bradley, B. (2006) 'Against Satisficing Consequentialism', *Utilitas*, 18/2: 97-108.
- Broad, C. D. (2013) 'Self and Others', in *Broad's Critical Essays in Moral Philosophy*, David R Cheney (ed.), Abingdon: Routledge: 262-82.
- Brown, C. (2017) 'Maximalism and the Structure of Acts', *Noûs*, Online First.
- Byron, M., ed. 2004. *Satisficing and Maximizing: Moral Theorists on Practical Reason*, Cambridge: Cambridge University Press.
- Clayton, M. and A. Williams (2000) *The Ideal of Equality*, Basingstoke New York: Palgrave ; St. Martin's Press.
- Curtis, B. (1981) 'The Supererogatory, the Foolish and the Morally Required', *The Journal of Value Inquiry*, 15/4: 311-18.
- Dorsey, D. (2012) 'Weak Anti-Rationalism and the Demands of Morality', *Noûs*, 46/1: 1-23.
- Frowe, H. (2014) *Defensive Killing*, Oxford: Oxford University Press.
- Gardner, J., et al. (2011) 'Desert and Avoidability in Self-Defense', *Ethics*, 122/1: 111-34.
- Gert, J. (2007) 'Normative Strength and the Balance of Reasons', *The Philosophical Review*, 116/4: 533-62.
- Hampton, J. (1993) 'Selflessness and the Loss of Self', *Social Philosophy and Policy*, 10/01: 135-65.
- Hedden, B. (2012) 'Options and the Subjective Ought', *Philosophical Studies*, 158: 343-60.
- Horgan, T. and M. Timmons (2010) 'Untying a Knot from the inside Out: Reflections on the "Paradox" of Supererogation', *Social Philosophy and Policy*, 27/02: 29-63.
- Horton, J. (2017) 'The All or Nothing Problem', *Journal of Philosophy*, 114/2: 94-104.
- Hurka, T. (2005) 'Proportionality in the Morality of War', *Philosophy & Public Affairs*, 33/1: 34-66.
- Hurka, T. and E. Shubert (2012) 'Permissions to Do Less Than the Best: A Moving Band', in *Oxford Studies in Normative Ethics, Volume 2*, Mark Timmons (ed.), Oxford: Oxford University Press: 1-27.
- Kagan, S. (1994) 'Defending Options', *Ethics*, 104/2: 333-51.
- Kamm, F. M. (1985) 'Supererogation and Obligation', *Journal of Philosophy*, 82/3: 118-38.
- Kamm, F. M. (1992) 'Review: Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status', *Philosophy and Public Affairs*, 21/4: 354-89.
- Kamm, F. M. (2007) *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*, New York: Oxford University Press.
- Kamm, F. M. (2011) *Ethics for Enemies*, Oxford: Oxford University Press.
- Lazar, S. (2013) 'Associative Duties and the Ethics of Killing in War', *Journal of Practical Ethics*, 1/1: 3-48.

- Lazar, S. (2016) 'The Justification of Associative Duties', *Journal of Moral Philosophy*, 13/1: 28-55.
- Lazar, S. (2017) 'Deontological Decision Theory and Agent-Centred Options', *Ethics*, 127/3: 579-609.
- Lazar, S. (2018) 'Moral Status and Agent-Centred Options', *Utilitas*, Forthcoming.
- McMahan, J. (1994) 'Self-Defense and the Problem of the Innocent Attacker', *Ethics*, 104/2: 252-90.
- McMahan, J. (2002) *The Ethics of Killing: Problems at the Margins of Life*, Oxford: Oxford University Press.
- McMahan, J. (2005) 'The Basis of Moral Liability to Defensive Killing', *Philosophical Issues*, 15/1: 386-405.
- McMahan, J. (2009) *Killing in War*, Oxford: Oxford University Press.
- McMahan, J. (2017) 'Doing Good and Doing the Best', in *Philanthropy and Philosophy: Putting Theory into Practice*, Paul Woodruff (ed.), New York: Oxford University Press.
- McNaughton, D. and P. Rawling (1991) 'Agent-Relativity and the Doing-Happening Distinction', *Philosophical Studies*, 63/2: 167-85.
- McNaughton, D. and P. Rawling (1995) 'Value and Agent-Relative Reasons', *Utilitas*, 7/1: 31-47.
- Mulgan, T. (2005) *The Demands of Consequentialism*, Oxford University Press.
- Otsuka, M. (1994) 'Killing the Innocent in Self-Defense', *Philosophy and Public Affairs*, 23/1: 74-94.
- Pettit, P. (1984) 'Satisficing Consequentialism', *Proceedings of the Aristotelian Society*, 58: 165-76.
- Portmore, D. W. (2011) *Commonsense Consequentialism: Wherein Morality Meets Rationality*, Oxford: Oxford University Press.
- Pummer, T. (2016) 'Whether and Where to Give', *Philosophy & Public Affairs*, 44/1: 77-95.
- Quong, J. (2009) 'Killing in Self-Defense', *Ethics*, 119/2: 507-37.
- Rogers, J. (2010) 'In Defense of a Version of Satisficing Consequentialism', *Utilitas*, 22/02: 198-221.
- Scheffler, S. (1994) *The Rejection of Consequentialism*, Oxford: Clarendon Press.
- Sider, T. (1993) 'Asymmetry and Self-Sacrifice', *Philosophical Studies*, 70/2: 117-32.
- Slote, M. (1984a) 'Satisficing Consequentialism', *Proceedings of the Aristotelian Society*, 58: 139-63.
- Slote, M. (1984b) 'Morality and Self-Other Asymmetry', *The Journal of Philosophy*, 81/4: 179-92.
- Slote, M. (1991) 'Review: Shelly Kagan's the Limits of Morality', *Philosophy and Phenomenological Research*, 51/4: 915-17.
- Southwood, N. (2011) 'The Moral/Conventional Distinction', *Mind*, 120/479: 761-802.
- Tadros, V. (2011) *The Ends of Harm: The Moral Foundations of Criminal Law*, Oxford: Oxford University Press.
- Vessel, J.-P. (2010) 'Supererogation for Utilitarianism', *American Philosophical Quarterly*, 47/4: 299-319.
- Wolf, S. (1982) 'Moral Saints', *Journal of Philosophy*, 79/8: 419-39.

---

<sup>1</sup> This paper stems from one that I presented at the Yale Moral and Political Philosophy Working Group and at the Kadish Center for Morality, Law and Public Affairs, at UC Berkeley. Many thanks to Yuan Yuan and Christopher Kutz for the respective invitations, and to the attendees for their careful comments and criticisms. Thanks to Dale Dorsey, Anne Gelling, Josef Holden, Tom Hurka and Doug Portmore for written comments. My work on this paper has been supported by Australian Research Council grant DP170101394.

<sup>2</sup> Paradigmatically: Kamm (2007).

<sup>3</sup> For example, Portmore (2011).

<sup>4</sup> Kamm (2007); McMahan (2002); Tadros (2011).

<sup>5</sup> See, for example, Lazar (2017).

<sup>6</sup> For an excellent overview of commonsense morality on this topic, see Hurka and Shubert (2012).

<sup>7</sup> See, for example, Scheffler (1994).

<sup>8</sup> Broad (2013); Slote (1984b).

<sup>9</sup> Hurka and Shubert (2012).

---

<sup>10</sup> Hampton (1993).

<sup>11</sup> Lazar (2018).

<sup>12</sup> Byron (2004); Slote (1984a).

<sup>13</sup> For example, in Bradley (2006); Mulgan (2005); Pettit (1984).

<sup>14</sup> Bradley (2006); Mulgan (2005).

<sup>15</sup> Interestingly, this objection is structurally parallel to Kagan's similar complaint against other arguments for moral options—that they threaten to overgenerate, resulting in options to inflict harm.

<sup>16</sup> Thanks to a referee here.

<sup>17</sup> For a lengthy discussion of the implausibility of requiring marginal interpersonal tradeoffs see Portmore (2011: Chapter 1).

<sup>18</sup> This possibility is either floated or defended in Curtis (1981); Dorsey (2012); Kagan (1994: 337); Portmore (2011); Slote (1991); Vessel (2010); Wolf (1982).

<sup>19</sup> DRAC is not Portmore's last word. His definitive principle is 'Commonsense Consequentialism', at Portmore (2011: 225). The difference between these principles is irrelevant for my purposes here, so it is much better to use the simpler one—which also has more in common with other rational pluralist approaches.

<sup>20</sup> 'Moral reasons either have some moral requiring strength or, if they do not, they are mere moral enticers. Moral enticers can make doing what they entice us to do supererogatory, but they cannot make doing what they entice us to do obligatory. Thus, a moral reason is a reason that, if sufficiently weighty, could make an act either obligatory or supererogatory. A reason that could only justify—that is, a reason that could not make an act obligatory or supererogatory but could only make an act permissible—would be a (morally relevant) non-moral reason.' Portmore (2011: 123).

<sup>21</sup> '[T]here is nothing, morally speaking, that counts in favor of promoting one's self-interest per se. This is not to say that one never has a moral reason to do what will further one's self-interest; one often does, as when doing one's moral duty coincides with promoting one's self-interest. The claim is only that the mere fact that performing some act would further one's self-interest does not *itself* constitute a moral reason to perform that act, for the mere fact that performing some act would be in one's self-interest is never by itself sufficient to make an act obligatory, or even supererogatory.' Portmore (2011: 128). See also Portmore (2011: 96, fn 39)

<sup>22</sup> Sider (1993); Vessel (2010).

<sup>23</sup> This last qualification caters for worries about the levelling down objection. See Clayton and Williams

---

(2000).

<sup>24</sup> For a popular view, see e.g. McMahan (2005).

<sup>25</sup> Witness, for example, the Crown vs. Dudley and Stephens case, in which sailors were tried for murder after eating the cabin boy, even though cannibalism was their only means of staying alive. This point is elegantly made in McMahan (1994); Otsuka (1994).

<sup>26</sup> Contemporary theorists of self-defence have not made this claim explicit in print, perhaps because it is presupposed by them all. But at least Jeff McMahan, Victor Tadros, Frances Kamm, Helen Frowe, and Jon Quong all share this view. See Frowe (2014); Kamm (2011); McMahan (2009); Quong (2009); Tadros (2011). And though it is poor evidence, it's worth also noting that no legal system in history has ever countenanced punishing those who fail to defend themselves against lethal threats (though some, of course, have treated attempted suicide as a crime).

<sup>27</sup> E.g. Frowe (2014); McMahan (2009); Tadros (2011). There are dissenters however; see for example Gardner et al. (2011).

<sup>28</sup> See, for example, Hurka (2005).

<sup>29</sup> Horton (2017); McMahan (2017); Pummer (2016).

<sup>30</sup> Pummer (2016).

<sup>31</sup> Christian Barry and I have a more developed view on this, in Barry and Lazar (2018).

<sup>32</sup> COST traces its ancestry to Scheffler (1994), though unlike Scheffler's hybrid views it is consistent with there being agent-centred restrictions; nor did Scheffler's theory address agent-sacrificing options.

<sup>33</sup> On these possibilities, see Brown (2017); Hedden (2012).

<sup>34</sup> See Lazar (2017).

<sup>35</sup> McNaughton and Rawling (1991, 1995).

<sup>36</sup> For one appealing account of what makes something a moral reason, see Southwood (2011). For the idea of using reactive attitudes in this way, thanks to Josef Holden.

<sup>37</sup> Horgan and Timmons (2010).

<sup>38</sup> Kamm (1992); Lazar (2018).

<sup>39</sup> For my take on what justifies associative duties, see Lazar (2016).

<sup>40</sup> This would be a similar verdict to that reached in the Crown vs. Dudley and Stephens case, mentioned above, in which the shipwrecked sailors who ate the comatose cabin boy were ultimately convicted of

---

murder, but sentenced leniently, ultimately serving only six months in prison. Though their action was clearly morally wrong, their exigent circumstances meant that it was substantially excused, so their punishment was accordingly lenient.

<sup>41</sup> Kamm (1985).

<sup>42</sup> Lazar (2013).

<sup>43</sup> Gert (2007).

<sup>44</sup> I think that the only published advocate of the production-costs view is Barry Curtis. But it does seem to be popular in discussion at least: 'if the cost or risk is considerably less significant [than the moral value of the end], the action is morally required; if considerably more significant, the action is foolish or unwise. But if the cost or risk is roughly *as* significant as the moral value of the end, the agent has done something which is "above and beyond the call of duty"—something which is morally good, but not morally required'. Curtis (1981: 311). For a more detailed discussion of this idea, see Barry and Lazar (2018).

<sup>45</sup> Hampton (1993); Hurka and Shubert (2012).

<sup>46</sup> Perhaps along the lines developed in Hurka and Shubert (2012).