MORAL SUNK COSTS[1]

Seth Lazar, ANU

Deontological ethics, nonconsequentialism, harm, sunk costs, sequential decision theory, dynamic choice.

Abstract

Suppose that you are trying to pursue a morally worthy goal, but cannot do so without incurring some moral costs. At the outset, you believed that achieving your goal was worth no more than a given moral cost. And suppose that, time having passed, you have wrought only harm and injustice, without advancing your cause. You can now reflect on whether to continue. Your goal is within reach. What's more, you believe you can achieve it by incurring—from this point forward—no more cost than it warranted at the outset. If you now succeed, the total cost will exceed the upper bound marked at the beginning. But the additional cost from this point is below that upper bound. And the good you will achieve is undiminished. How do the moral costs you have already inflicted bear upon your decision now?

1. Introduction

Sometimes the world cooperates with our plans. We can achieve our goals, or vindicate our rights and others', without incurring any moral costs along the way. Justice and the good come for free, with no need to inflict or

[1] I presented this paper at MIT and UNC Chapel Hill. Thanks to the audiences at those talks, and to the conveners for inviting me. Thanks also to the following for comments and helpful discussion: Christian Barry, Anne Gelling, Caspar Hare, Jeff McMahan, Emily McTernan and Victor Tadros. Work on this paper was supported by ARC grant number DP170101394.

endure harms or injustices in their pursuit.

More commonly, however, the world is not so compliant. We have to fight for the good, fight for our rights. And casualties are inevitable. We bear costs ourselves, and we inflict them upon others. Pursuing our ends can be right only if they are worth those costs. So how do we determine whether our moral goals are worth the moral costs?

Recent work in the ethics of war has revealed a new puzzle within this most basic of moral questions.[2] Suppose that, at the outset, you believed that achieving your goal was worth incurring a given amount of moral cost; if you knew you would incur more, it would be wrong to proceed. And suppose that, time having passed, you have wrought only harm and injustice, without advancing your cause at all. You can now reflect again on whether to continue. Your goal is within reach. What's more, you believe you can achieve it by incurring—from this point forward—no more cost than it warranted at the outset. If you now succeed, the total cost will exceed the upper bound marked at the beginning. But the additional cost from this point is lower than that upper bound. And the good you will achieve is undiminished. How do the moral costs you have already inflicted bear upon your decision now? Should you treat them as economists think we should treat 'sunk costs', and disregard them?[3] Or should they somehow affect

---

[2] Fabre (2015); Kamm (2001); McMahan (2015); Moellendorf (2015); Rodin (2008, 2015).

[3] For a discussion of the 'sunk cost fallacy' in rationality, see Kelly (2004).

what you are permitted to do now?

This kind of case readily arises in the use of force. In the terms of just war theory, our topic is proportionality. A war of defence against aggression, for example, might be proportionate at T_1, provided its goals can be achieved without inflicting more than X civilian casualties. At T_2, we might find ourselves with X civilian casualties already, but without having reversed the initial aggression. We must then ask whether it is proportionate to continue fighting, given that we will exceed our initial 'proportionality budget'.[4] This quandary can also arise for particular actions within war, as well as in uses of force outside war. Suppose, for example, that an initial armed attempt to free some hostages has failed, and the number of innocent casualties already exceeds what would, at the outset, have been proportionate to ending the crisis. How should those 'moral sunk costs' affect what we ought to do now?

Questions of proportionality also arise in other contexts: for example, in risky rescue attempts, emergency medicine or the management of bushfires. And the problem arises in more mundane scenarios too—environmental regulation and public health policy, for example; even taxation. In all these areas, you want to achieve some moral goal, doing so is 'worth' a particular degree of moral cost and no more; at some point during the process, you reassess and realise that you can no longer achieve the objective

[4] Moellendorf (2015).

and remain within your antecedent 'proportionality budget', but if you look only forward, the *remaining* costs do seem justified by the *remaining* good.

The nascent discussion of moral sunk costs has settled into two camps. Adopting Jeff McMahan's terminology, we have on one side the 'Quota View', and on the other the 'Prospective View'. Adherents of the Quota View think the initial proportionality budget is absolutely constraining.[5] 'Prospectivists' think that at each decision point we should look only forward: sunk costs are irrelevant, except insofar as they constitute evidence against our more optimistic projections about how things will go from now on.[6]

I think that neither side has captured the whole truth: I propose an intermediate approach, which I call the Discount View. Of course, a mere compound of two positions might deliver the right verdicts on cases, but still be explanatorily inadequate. I show that the Discount View is grounded in a sensible underlying moral theory, and offer some arguments in its favour. I then consider some implications and objections, before concluding.

2. Framing the Problem

Moral sunk costs raise many interesting conceptual and normative issues.

---

[5] Fabre (2015); Moellendorf (2015).

[6] McMahan (2015); Rodin (2008).

My focus in this article is comparatively narrow. I want to ask whether and how moral sunk costs can render an act disproportionate. There are many different interpretations of the proportionality condition, and I cannot consider them all.[7] So, for my purposes, an act is disproportionate if and only if its expected moral costs outweigh its expected moral benefits.

An act's moral costs are all the negative considerations that it brings about, as determined (and weighted) by one's broader moral theory. That theory might give more weight to costs of which the act is a more proximate cause, or to costs that are intended rather than merely foreseen, and so on. Benefits are the positive analogue of costs, and are subject to similar weightings. By the 'expected costs' of an act I mean the probability-weighted average of the costs in the possible outcomes of that act. To determine the expected costs for an act ϕ, multiply the cost of each possible outcome of ϕ-ing by that outcome's probability of coming about; the sum of those products is the expected cost. Work out the expected benefits in a parallel way. Determining which probabilities count is itself a deeply vexed question. To fix matters, I stipulate that the relevant probabilities are grounded in the evidence available if one does the morally appropriate research.

Mine, then, is an 'evidence-relative' understanding of proportionality. We could also understand proportionality as 'fact-relative' or 'belief-

[7] The canonical take is Hurka (2005). For a more recent view, see McMahan (2018).

relative', among other possibilities.[8] Most philosophers understand proportionality as necessary but not sufficient for permissibility. But we can understand permissibility in these different ways as well, so each interpretation of proportionality must be linked to an interpretation of permissibility. The most plausible approach keeps both standards in the same epistemic register: an act's being evidence-relative proportionate is a necessary condition of its being evidence-relative permissible; an act's being fact-relative proportionate is a necessary condition of its being fact-relative permissible, and so on. Mixed alternatives are possible, but I will not consider them further. The ensuing discussion will focus on evidence-relative proportionality, and evidence-relative permissibility.

With these stipulations in place, we can frame the central question of the paper more precisely. Suppose you want to pursue a morally worthy goal, G. Your alternatives at T_1 are to ϕ or to do nothing. ϕ-ing has some prospect of realising G. Its expected benefits are B, its expected costs C. C is equal to B, so at T_1 ϕ-ing in pursuit of G is *just* proportionate.

So, you ϕ, but things don't go according to plan. By T_2 you have achieved nothing of value, but you have already inflicted an actual cost equal to C. But now you face a choice whether to ψ or to do nothing, where ψ-ing has some prospect of realising G. The expected costs of ψ-ing are C, and the

[8] Parfit (2011).

expected benefits are B—just the same as for ϕ-ing. Should the costs incurred when you ϕ-d have any bearing on whether it is proportionate for you to ψ?

There are three possible responses. First: costs incurred at T_1 increase the urgency of realising G at T_2, and hence the benefit of doing so. One might think, for example, that we have additional reason at T_2 to redeem the sacrifices made following the decision at T_1.[9] Although this clearly chimes with some common intuitions, I want to focus instead on whether moral sunk costs incurred by ϕ-ing can count *against* the proportionality of ψ-ing.

The Prospective View says that they cannot. The costs incurred by ϕ-ing are in the past; you can do nothing about them, so they are irrelevant to whether ψ-ing is proportionate. If ϕ-ing was proportionate, given C and B, then so is ψ-ing, since C and B here are projected to be the same as before.

By contrast, on the Quota View, the pursuit of G gives us a fixed proportionality budget. Once that budget is exhausted, it is disproportionate to incur any further moral costs to that end. The costs incurred when you ϕ-d at T_1 used up your proportionality budget for the pursuit of G, so it is disproportionate to ψ at T_2.

The Discount View combines the best features of the Quota and

[9] McMahan (2015).

Prospective Views. But before introducing it, it will help to offer an example to illustrate the difference between those alternatives and highlight the need for a new approach.

3. Shortcomings of the Quota and Prospective Views

The problem of moral sunk costs pervades many of our weightiest and most morally serious endeavours. It is a complex problem: to clearly identify the difference between the competing views, we need a very simple case, without the complicating distractions involved in any more realistic scenario.

> *Iterated Loop*: A trolley is heading towards five innocent victims, who can be saved only if you divert it. It is approaching a junction, controlled by a probabilistic lever. If you pull the lever, then there is some probability, *p*, that the trolley will head down the track called STOP, where it will kill nobody, and come to a halt. But there is some probability, 1-*p*, that it will instead head down the LOOP track, where it will kill one person, and then loop round to the start, again heading towards the five. The LOOP victim will immediately be replaced, leaving you with the same decision at T_2 as you faced at T_1, with just the same odds; the same holds for T_3–T_n.

All the potential beneficiaries and victims are there through no fault of

their own; each has as much to live for as the others. The lever is governed by quantum mechanics, so its probabilities are true objective chances, known in advance.

Suppose that at T_1 you pull the lever. You are unlucky; the trolley heads down LOOP, killing one innocent person. How, if at all, should the death already inflicted in the attempt to save the five affect what you ought to do at T_2, when the trolley approaches the junction for a second time, and you must decide again whether to pull the lever?

Let's use the ratio familiar from most contemporary discussions of trolley cases, and stipulate that for an expected benefit equivalent to five lives saved, any more than one expected death inflicted would be disproportionate. You have already inflicted one death in the attempt to save the five, so the expected total cost of pulling the lever at T_2 having pulled it at T_1 is one certain death plus 1-*p* chance of another death. The Quota View, then, must say at T_2 that it is disproportionate to pull the lever.

The Prospective View says that costs inflicted at T_1 are irrelevant to what you ought to do at T_2. The expected costs and benefits that determine whether it is proportionate to proceed at T_2 are exclusively those that have not yet occurred. It follows that your decision at T_n is the same as it is at T_1. If it was proportionate to pull the lever at T_1, then it must be proportionate to do so at T_n. Even if the trolley were to keep looping ad infinitum, if you should pull the lever at T_1, then you should do so every time subsequently.

Framed in this way, it is easy to see why neither of these views is wholly satisfactory. The Quota View gives us a strict proportionality budget to use in the pursuit of any given good. Once that budget is used up, no further risks may be run, no matter how much is at stake. Jeff McMahan has argued that it would be absurd to turn down the opportunity to certainly achieve your goal at T_2, just because doing so would involve some comparatively negligible moral cost.[10]

Consider a variant on the above case—call it *Iterated Loop (Finger)*. Suppose that *p*(STOP) = 0.5. At T_1 you pulled the lever, but were unlucky: the trolley went down LOOP, killing one person. Now at T_2, the trolley is coming around for the second time. But this time, the person on LOOP is at risk of a much lesser harm—losing a finger, say. Had you known that all this would happen from the start, then it would have been disproportionate to pull the lever at T_1—the expected costs would have included one death for sure, and 0.5 probability of a finger, which by hypothesis exceeds what can be justified by the prospect of saving the five. So the Quota View has to say that you must stop. You cannot pull the lever again. Your reasons to save the five cannot justify any further costs on their behalf.

In this case, the Prospective View gives the intuitively plausible verdict, because it disregards moral sunk costs. But Moellendorf has argued that

[10] McMahan (2015: 702). Victor Tadros has also been a prominent proponent of this objection, in forthcoming work.

the Quota View can get this case right.[11] In cases like these, he suggests, we should reassess our initial proportionality calculation, and argue that the benefit of realising the goal is in fact worth one death and a 0.5 probability of cutting off a finger. We might bolster his view by arguing that the relata in the proportionality calculation are inherently vague, so the boundary between what counts as proportionate and disproportionate is fuzzy enough to subsume small additional expected costs.[12]

Although these responses soften the implications of McMahan's objection for the Quota View, they do not offer a decisive solution. Suppose that we revise the proportionality budget, as Moellendorf suggests, allowing you to pull the lever at T_2, in *Finger*. But you are again unlucky, causing the person on LOOP to lose his finger. The trolley returns to the beginning, and another victim is placed on LOOP, her finger at stake: the total expected cost of pulling the lever at T_3 is now a life and one finger, plus 0.5 probability of another finger being lost. If we again revisit the initial proportionality calculation, arguing that saving the five is worth that much expected cost, then we can simply ratchet things up again. We can keep doing this until it is simply implausible to assert that the counterintuitive implication can be explained away by appealing either to an initially incorrect proportionality calculation or to the vagueness of the comparisands. It will always be

[11] Moellendorf (2015: 668).

[12] Thanks to Emily McTernan for this point.

counterintuitive to refuse the opportunity to save five lives at the cost of a finger, no matter how many fingers have already been lost.

Consider, for example, *Iterated Loop (0.999)*, in which *p*(STOP) = 0.999. Suppose that you pull the lever at T_1, and the trolley goes down LOOP, killing an innocent victim. At T_2, according to the Quota View, pulling the lever has an expected total cost of 1.001 expected deaths—the one death inflicted at T_1, and 0.001 expected deaths from pulling the lever at T_2. Suppose, as Moellendorf argues, that this is close enough to one expected death that pulling the lever still counts as proportionate. Again you are unlucky, and are faced with the same decision. If you pull the lever at T_3, then the expected cost is 2.001 expected deaths—the deaths inflicted at T_1 and T_2, and a further 0.001 prospect of taking the third victim's life on LOOP. That amount of expected cost would clearly be disproportionate even to a certainty of saving the five. So it is impermissible to pull the lever at T_3. Indeed, it would be impermissible even if the probability of the trolley going down STOP, killing nobody, was arbitrarily close to 1. This is an unavoidable implication of the Quota View, and is deeply counterintuitive.

The Prospective View fares much better in these cases. Because it disregards moral sunk costs, if the expected benefits are worth the expected costs from this point forward, then it is proportionate to proceed, regardless of costs already incurred. But this leads to its own problems. As Rodin, Fabre and Moellendorf have all argued, there is in principle no stopping point

to the ratcheting up of moral costs incurred in the pursuit of some finitely valuable objective.[13] While it seems plausible that we should always be able to run a small risk for the sake of a great benefit, no matter how much cost has already been incurred, the Prospective View is much more extreme than this.

Return to *Iterated Loop*. Suppose that p(STOP) = 0.5. And suppose that there is an infinite supply of victims who can be placed on LOOP (imagine the case to be a supertask, in which an infinite series of actions can be performed in a finite time). Call this variant *Infinite Loop*. Assuming the standard 1:5 moral mathematics, the Prospective View should say that it is proportionate to proceed at T_1, and indeed at *every* iteration of the problem, ad infinitum: the expected cost of pulling the lever at every iteration is one expected death caused, the expected benefit is five lives saved. More generally, if it is proportionate to proceed at T_1, then it is proportionate to proceed at T_n for any n. So the Prospective View implies that it can be proportionate to countenance inflicting infinite costs for the sake of realising a finite good, with constant stakes throughout. This seems like no less a theoretical cost than that faced by the Quota View.

It is important not to overstate this point. Moellendorf's claim that no

[13] Fabre (2015: 663-5); Moellendorf (2015: 637), Rodin (2008: 58) thinks that the Prospective View's adherents should bite the bullet.

finite good is worth infinite costs is true, but not apposite.[14] Our question is whether the *expected* good is worth the *expected* cost. And the Prospective View does not imply that a finite expected good can outweigh infinite expected cost. Even though in the infinite variant of *Iterated Loop* it is possible for the same decision problem to be repeated an infinite number of times, the probabilities decrease as the possible costs increase, so the expected cost remains finite.[15] But still, how can it be proportionate to even *entertain* incurring infinite costs for the sake of a finite good, continuing at every iteration to run the same risk of things going wrong? One can't help but feel the pull of Rodin's, Fabre's and Moellendorf's worry that, on the Prospective View, the proportionality constraint has too little bite.

4. The Discount View

Each side of this debate has made a strong case against the other. The Quota View avoids being excessively permissive at the cost of being excessively strict. The Prospective View makes the reverse trade-off. Each view is consonant with only half of the 'intuitive data'. The obvious solution is to find a middle way between them.

To improve on the Prospective and Quota views, our principle must permit incurring small risks for the sake of significant expected benefits,

[14] Moellendorf (2015: 665).

[15] See Lazar (2016).

but prohibit indefinitely and repeatedly running the same magnitude of risk for the sake of a constant probability of achieving the same goal.. The solution: as you rack up costs in the failed attempt to pursue a morally worthy objective, your reasons to secure that objective progressively diminish in weight (are discounted), until they reach a lower bound beneath which they cannot drop. In other words, you always have *some* reason to realise your objective. But your reasons diminish in weight as you incur more costs for that objective's sake, asymptotically approaching a lower limit.

A principle like this would let us have our cake and eat it too. We could adopt the permissiveness of the Prospective View in cases like *Finger* and *0.999*: no matter how many times the lever has been pulled, the expected benefit of saving the five is always worth enough to justify an additional risk of someone losing their finger, or a 0.001 risk of killing an innocent person—and could arguably justify more than that. But we can channel the Quota View in other cases: the expected benefit of pulling the lever at T_2 is less than that at T_1, and in general is less at T_{n+1} than at T_n, so at some point in cases like *Infinite Loop*, where the risks are held constant, it must become disproportionate to proceed. This means we need not countenance endlessly imposing the same magnitude of risk for the sake of a constant probability of achieving the same goal.

Neither the Quota nor the Prospective View has been given firm theoretical foundations. Both rest, thus far, only on intuitions about cases. On that

score, the Discount View is already in good shape. Some cases favour the idea that past costs can render it impermissible to keep imposing significant risks for the sake of the same benefit; some cases favour ignoring past costs in deliberations about proportionality. We can accommodate all of those intuitions in a view that our reasons to act in iterated decision problems can diminish in force asymptotically to some lower bound, so that they always justify imposing *some* additional risk of harm, but cannot, at T_n, justify imposing as much risk as they would justify at T_1.

However, I do not want to rely only on intuitions about cases to defend the Discount View. I want to make a theoretical case for it—while acknowledging that, given the diversity of views on the underlying moral theory, it is good to remain as ecumenical as possible.

To motivate the Discount View, I need only the idea that, when others' well-being is at stake, our reasons to help them are grounded in one of at least two facts. First, their well-being is intrinsically valuable. This grounds reasons to promote their well-being. Second, *they* are intrinsically valuable: they have moral status, and matter independently of how their well-being contributes to the world. This grounds reasons to show them appropriate respect.

I will write that we have *well-being-based reasons*, grounded in the intrinsic value of well-being, and *status-based reasons* grounded in equal moral status. This may be somewhat factitious, since one could say that all of our moral reasons are grounded in both sources. I will return to that possibility

in Section 5. For now, it helps keep things clear to consider them as two kinds of reasons.

Our status-based reasons are grounded in the importance of equal respect for those of equal moral status, and are most notably associated with our rights, both to be helped, and not to be harmed. Typically, failure to act on a status-based reason towards another person amounts to a failure of respect, and, at least pro tanto, wrongs the victim. Of course, status-based reasons need not be verdictive: sometimes we are all-things-considered justified in committing an act that *pro tanto* wrongs a person.

Our well-being-based reasons are reasons to avert suffering and bring about happiness, just because suffering is bad and happiness is good. Failure to act on a well-being-based reason might be wrong, but does not wrong anyone in particular, nor does it involve a failure of respect.

Some old-fashioned utilitarians might believe that well-being-based reasons exhaust the moral domain. In other words, we *only* have reasons to promote well-being.[16] Some hard-line deontologists will think that we *only* have status-based reasons, and have no reason to promote well-being as such. The Discount View is most interesting if these extreme views are false, but if you're an old-fashioned utilitarian, then I think my arguments show that you should endorse the Prospective View, while if you're a hard-line deontologist you can still endorse the Discount View (see Section 5).

---

[16] They don't have to, though: for an example of a utilitarian theory that can accommodate the distinction between status-based and well-being-based reasons, see Chappell (2015).

It is worth pausing to consider one objection that might come from deontologists at this point. They might agree on the distinction between status-based reasons and well-being-based reasons, but deny that it is relevant to trolley cases, arguing that that we have *only* well-being-based reasons to turn the trolley; and we have *only* status-based reasons not to do so.[17] I have more and less concessive responses.

The first response is concessive. Even if this interpretation of trolley cases is true, the Discount View might still be right. Perhaps it does not apply in trolley cases, but does apply to others where status-based reasons are active. Or perhaps our well-being-based reasons, properly understood, do in fact diminish in the way needed to support the Discount View (more on this in Section 5).

A less concessive—and I think correct—response: status-based reasons *are* in play in standard trolley cases; they *can* tell both for saving the five, and against killing the one. Two arguments:

First, suppose that the five were culpably responsible for the one being on the loop track, in *Infinite Loop*. In that case, there would be no status-based reason to save them. They would be liable to bear their deaths, to ensure that their victim is not killed, so would not be wronged by being left to die. Still, you would have *some* well-being-based reason to help them. If you could save them by turning the trolley towards the one, knowing that

[17] Thanks to a referee for pushing me on this point.

he would suffer only an injured foot, then it might be permissible to do so. But this is quite different from the proportionality ratio in standard trolley cases. When the five lack rights to life, the proportionality ratio radically changes. This suggests that in ordinary trolley cases (without culpability), the five *have* rights to life. Since reasons to protect and preserve people's rights *are* status-based reasons, this implies that in standard trolley cases, you have status-based reasons to save the five.

The second argument shows that there clearly are status-based reasons in some non-standard trolley cases, which implies that they are also present in standard cases. Suppose you can save the five by turning the trolley down an empty side-track, harming nobody, at negligible personal cost. If you let the five die, you seriously wrong them. This would be not only a gratuitous failure to realise some good, but a failure of respect.

Now suppose that there is someone on the side-track, who will die if the trolley is turned. If the trolley is not turned, however, 1,000 people will die (raise the number if you like).[18] Now you seem clearly required to turn the trolley. Not turning it would wrong the 1,000. This implies that you have status-based reasons to aid the 1,000.[19] It does not entail, of course, that you

[18] Helen Frowe argues that even in 5:1 cases you are morally required to turn the trolley, her account of our duties of rescue, in my terminology, amounts to arguing that you have status-based reasons to turn the trolley in such cases. Frowe (2018).

[19] I say 'implies', because it is possible that the 1,000 might be wronged, without status-based reasons being engaged, in which case this argument would not go through. However, I think that someone having a justified complaint is pretty good evidence that a status-based reason has been contravened.

have *only* status-based reasons to aid the 1,000.

If we have status-based reasons to aid the five in the no-cost case, and status-based reasons to aid the 1,000 in the last case, it is highly likely that we have status-based reasons to aid the five in standard trolley cases. The most plausible explanation is that in each of these variations your reasons to save any given person are the same: the difference between the cases just depends on (1) how many other interests are aligned with that individual's interest, and (2) the aggregate weight of the competing reasons.[20]

One could argue that we can only determine which reasons tell in favour of saving a person when we hold fixed the reasons that tell against it.[21] If this is right, then we cannot infer that status-based reasons are at work in the 5 v 1 case from their being at work in the 5 v 0 and 1,000 v 1 case. However, I don't think that it is generally true that we can work out which reasons tell *for* a particular action, only by first considering which reasons tell *against*. It is much more plausible that your reasons to save any particular person remain constant across these cases, with the only difference being how many other people you can save, and whether someone has to be sacrificed.

So, we have reasons to aid others, grounded in both the intrinsic value of

---

[20] As a referee points out, this means that status-based reasons can sometimes be aggregated. This is consistent, however, with aggregation being barred in some cases—e.g. where we weigh saving one life against averting some very large number of headaches.
[21] Thanks to a referee for this point.

their well-being, and in their moral status. When assessing whether an action is proportionate, we must weigh these reasons together. My proposal: our status-based reasons weaken as we incur more moral costs in the failed attempt to achieve our goals. Our well-being-based reasons, however, retain their full force.

This proposal raises many questions. First: why should we think that our status-based reasons and our well-being-based reasons differ in this way?

Our well-being-based reasons are grounded in the intrinsic value of others' well-being. The intrinsic quality of that well-being is unaffected by the moral sunk costs incurred in the failed attempts to realise it. Of course, contingently, the beneficiaries' lives might be less happy, tainted by regret for the costs incurred in the effort to save them. But we can stipulate this away by saying that everyone has just as much to live for as everyone else. Whether you save the five at T_1 or T_{10}, the welfare value that you realise by doing so is the same. So, reasons grounded in the intrinsic value of well-being will persist undiminished. What's more, since saving five lives is always a good thing to do, we will always have sufficient reason in variations on *Iterated Loop* to justify running small risks of serious harms, or high risks of less serious harms.

This point guarantees that our reasons to save the five will always remain above some lower bound. This definitively separates the Discount View from the Quota View, which asserts that at some point no further

risks can be justified. It also lends support to the Prospective View for those who think that we have only well-being-based reasons to help the five in trolley cases. To part ways with the Prospective View, I now need to show that our status-based reasons *do* diminish in force from one iteration to the next. I have a number of arguments to that effect.

First, though, note that I want only to show that the weight of our status-based reasons to achieve a goal G can be diminished by the costs inflicted in failed attempts to realise G. I take no stand on whether our status-based reasons can be wholly exhausted (I think they cannot), nor on precisely what the discount rate should be (I think it is unlikely to be very steep). I want only to show that there is some such discount rate. I have two kinds of arguments: the first appeals to specific features of status-based reasons; the second presents variations on *Iterated Loop* that favour the Discount View.

I have four arguments in the first category. The first is the most general. If you have a status-based reason to help me, then, in general, *at least pro tanto*, you owe it to me to help me, and if you fail to do so, you wrong me (*pro tanto*). Suppose that, in *Iterated Loop*, the probability of the trolley going down LOOP is 0.01. Failing to pull the lever at T_1 would, I think, wrong the five. But suppose that you pull the lever at T_1 and T_2, and twice you get unlucky, killing two innocent people. Is it at all plausible that, at T_3, failure to pull the lever would wrong the five in just the same way as failure to do so at T_1 would have? Is it plausible that you *owe* it to the five to pull the lever

at T_3, that failing to pull it would show them disrespect? I think not. You have already shown your respect for the five by pulling the lever twice, taking a serious moral risk on their behalf, and causing two innocent deaths. It is much less plausible that you owe it to them to pull the lever at T_3, having already incurred those costs. Of course, it might still be the right thing to do, because of the great good you can achieve by pulling the lever, but that would have to do with your well-being-based reasons to save the five, more than your status-based reasons.

The second argument draws on ideas of fairness. Your reasons to save the five in *Iterated Loop* may be partly grounded in the importance of giving them a fair chance of survival. [22] Suppose that *p*(LOOP) is 0.5. You pull the lever at T_1, and the trolley goes down LOOP, killing one. You face the same decision at T_2. But the argument that you now owe the five a fair chance of survival carries less weight. Even though it did not turn out well, they have at least had *some* chance of survival. Indeed, at some point the person on LOOP might be able to argue that you have already given the five a greater chance of survival than you would give him by pulling the lever, so he has a claim, grounded in fairness, that you not pull the lever again.[23]

Could one reply that, since you now know that the trolley went down

[22] See Rasmussen (2012); Taurek (1977); Walden (2014). Thanks here to Christian Barry.

[23] Thanks to Christian Barry for this point.

LOOP, you did not give the five *any* chance of survival at T_1?[24] Since I stipulated that the lever was governed by objective chance, this response is somewhat illicit. Nonetheless, in realistic cases the probabilities are likely to be only epistemic, and in some of these a failed attempt will reveal that the epistemic probability was misleading. And yet, even if the flip of a coin (say) is fully determined, it is still an intuitively fair way to resolve a dispute between parties with an equal claim. We have a claim that others *treat us* fairly. And how they treat us depends on what their evidence is. So giving someone a fair chance simply means giving them a reasonable prospect on your evidence. So, at T_2 you *have* already given the five a fair chance of survival, so this should diminish their claim to aid.

The third argument draws on the closely-related consideration of reciprocity. The Prospective View says that we can rack up sacrifices for the sake of the intended beneficiaries indefinitely, far past their actual capacity to ever reciprocate. Now, of course some of our rights are independent of our ability either to reciprocate or to 'pay it forward', but still, our ability to reciprocate sacrifices made on our behalf surely gives them some additional weight.

Return to *Iterated Loop*. Suppose you have pulled the lever five times, killing five victims on LOOP. Now the trolley approaches the junction for the sixth time. If you pull the lever and it again takes LOOP, then the five

---

[24] See e.g. Wasserman (1996). Thanks to Victor Tadros for raising this objection.

cannot, even in principle, sacrifice for others to the same degree as others have already sacrificed for them—they have only five lives to give. Their claim that others *continue* to bear risks for their sake is at least somewhat diminished. Our duties of rescue are not grounded in reciprocity alone, and some beneficiaries will forever be in 'moral deficit'. But these duties at least have a *dimension* that rests on the idea that we could all, in principle, be the ones to bear those costs for others' sakes. And if we completely disregard moral sunk costs, then situations can arise in which the beneficiaries of our risky rescue attempts take more from others than they could ever realistically endure for others' sakes.

This leads to the fourth argument. We typically think that people whose rights are infringed in the all-things-considered permissible attempt to preserve or vindicate another person's rights are entitled to compensation for their losses. When Feinberg's backpacker breaks into the hunting lodge, he clearly owes its owner compensation for the damage.[25] Civilians whose property is destroyed as a foreseeable 'collateral' harm in a just war are, in principle, entitled to compensation. Again, in principle, the most natural people to bear the cost of that compensation are those for whose sake the costs were incurred. And if we have no regard to sunk costs at all, then those compensatory obligations will mount beyond the point where anybody could possibly address them. By continuing to pull the lever, we are

[25] Feinberg (1978).

saddling the would-be beneficiaries with a debt that they cannot pay.

Considerations of respect, fairness, reciprocity and compensation all suggest that the status-based reasons to aid the intended beneficiaries of your action diminish in weight as you try and fail to save them, imposing costs on others along the way. None of these arguments, however, bears on the well-being-based reasons that you have to save the five. Whatever else is true, saving five innocent happy lives is always a good thing, that you have some reason to bring about. The magnitude of the well-being realised if you save the five is unaffected by issues of respect, fairness, reciprocity and compensation.

Notice, also, that these arguments do not entail any particular discount rate, or any particular view on whether our status-based reasons can be fully exhausted. They all simply give *some* justification for the status-based reasons being discounted at one iteration relative to the one before it.

The rest of my case for the Discount View rests on some further variations on *Iterated Loop*.

Suppose, first, that after you pull the lever at T_1 and the trolley goes down LOOP, the five victims are changed. You had no knowledge that this would happen, and no way of finding out. But now, at T_2, you have killed one person, and five different people are at risk. What you did to save the five at T_1 has no bearing on your reasons to save the five at T_2. So from the perspective of your status-based reasons, your decision at T_2 should be

identical to your decision at T_1 (no discount). Meanwhile, the well-being-based considerations are just the same at T_2 as they were at T_1—different bearers of well-being, to be sure, but the same amount. So, again, the case for pulling the lever at T_2 is no weaker than it was at T_1. Indeed, the fact that you are somewhat responsible for the second group of five being on the track might affect what you ought to do at T_2—perhaps your status-based reasons to help them are stronger than would otherwise be the case. This supports the Discount View, which predicts that the sunk costs worry arises from a discount in our status-based reasons, which is absent from this case.

Next, suppose that you are faced with two identical track setups. The probabilities are the same, but the second has only just started running, while the first has been running for three unsuccessful pulls of the lever. You must choose between the two tracks—you cannot pull both levers. I think you have stronger reasons to pull the lever on the second track. Three people have already lost their lives for the sake of the five on the first track; they have already had a fair chance at survival; the five on the second track have a clean slate, and have had no chance at survival yet, and so have a stronger claim to aid. If you share this intuition, then that also favours the Discount View (though it is also consistent with the Quota View; it tells against the Prospective View).

Third, suppose the same person is on the LOOP track for every iteration of *Iterated Loop*, and instead of death being at stake, anyone hit by the trolley will suffer searing but temporary pain that is soon forgotten. After a

point it is clearly impermissible to switch the trolley—the five must take the hit. There is a clear limit to what they can expect one person to bear on their behalf. This suggests there should also be a limit to what they can expect some larger number of people to endure for their sake. Of course, there is a difference between spreading costs out over many people, and concentrating them all on one person. Nonetheless, the lesson from this case is that aggregate costs matter. This supports the Discount View and the Quota View, but looks more troublesome for the Prospective View.

Finally, suppose that you are one of the five, and you can choose whether to pull the lever. If you have a claim to be saved, grounded in status-based reasons, then you can permissibly enforce that claim by pulling the lever. Suppose you have pulled the lever twice already, and killed two people. Could you permissibly pull the lever a third time? Do you have a claim, grounded in considerations of equal moral status, to do so? I think not. Of course, this just restates the conclusion that I am trying to argue for, but certainly I feel its intuitive pull even more strongly when I imagine being both the beneficiary and the person inflicting the sunk costs.

Together with the arguments from respect, fairness, reciprocity and compensation, as well as the Discount View's ability to avoid the counter-intuitive implications of both the Quota View and the Prospective View, and its reliance on a plausible picture of our underlying reasons to aid others, these additional intuitively plausible results lend the Discount View further support. In the next section, I consider some of the view's

implications, and some objections to it.

5. Implications and Objections

The first implication of the Discount View is that, in choices potentially involving moral sunk costs, we must attend to the kinds of reasons at stake. If the choice is driven by the importance of promoting well-being, then we should lean towards disregarding sunk costs. If it is driven by respect and status considerations, then we should be more inclined to take sunk costs into account. And in this latter case, we must also attend to whether the costs were incurred for the sake of the particular people who would benefit from continuing one's pursuit of the objective at stake.

Warfare is the paradigm case in which moral sunk costs matter. The prosecution of war *always* involves violating fundamental rights. So it can be proportionate only if done in the pursuit of our most fundamental individual and collective rights. If status-based reasons are not at stake, then there is no chance the war is proportionate. We don't fight wars to promote well-being. What's more, in most conflicts (though perhaps not the most protracted), the population in whose defence the war is fought is relatively stable. So moral sunk costs must be taken into account, and we must be sensitive to the costs already inflicted as we consider whether it is proportionate at T_2 to continue a war that was justly begun at T_1.

Few areas of public policy are exclusively aimed at the promotion of well-being; ordinarily rights are at stake, one way or the other. But in long-

term policy decisions there may often be population turnover—this is likely to be true for environmental policy for example—which would imply that sunk costs can more often be disregarded in that area.

The second implication: individuals must anticipate moral sunk costs at the outset of a potentially iterated decision problem. I discuss this in detail elsewhere, but here I want to emphasise two points.[26]

(a) If there is some probability that you will achieve your goal only at T_4, say, rather than at T_1, then the good that you will realise will be lesser, in proportion to the degree of discount that would apply to the claims of the beneficiaries given the three failed attempts to save them. This, in turn, makes pulling the lever at T_1 somewhat harder to justify. It raises the bar needed for pulling the lever to be proportionate. The effect will resemble David Rodin's 'moral contingency' in the proportionality budget—taking the Discount View into consideration will mean that at T_1 some options which would be proportionate under the Prospective View will be disproportionate.[27]

(b) Additionally, if we take 'future sunk costs' into account from the outset, as the Discount View says that we must, then we know that, in an iterated decision problem, there may come a time when we should stop. So we have to ask ourselves, at T_1, whether we will give up if that time comes. The

[26] [omitted].

[27] Rodin (2015).

'fallacy', recall, is to view sunk costs as a positive reason to continue the pursuit of one's goal. Many of us are susceptible to it. And that can be a reason against starting the process.[28]

So, the Discount View seems to capture the intuitive data of the Prospective and the Quota Views, without the shortcomings of either. It is grounded in a plausible picture of our underlying moral reasons. And it is well-motivated: the intrinsic value of well-being does not diminish as you rack up moral costs in its service; but the weight assigned to status-based reasons plausibly does. There are, however, several objections to address.

Any intermediate position is likely to take fire from both sides. The first pair of objections are likely to be raised by Prospectivists; the second pair by those who favour the Quota View.

The first objection involves questioning the intuitive foundations of the Discount View. McMahan argues that we would feel intuitively compelled to stop at some point, in cases like *Infinite Loop*, simply because in any realistic scenario, our continuing bad luck would give us more reason to doubt that the next attempt will succeed.[29] We also know in advance that in risky activities we will systematically miscalculate the probabilities—think of the 'gambler's fallacy', or the general belief that if you keep plugging away,

---

[28] This involves taking a stance in the debate between actualists and possibilists (see e.g. Jackson and Pargetter (1986)).
[29] McMahan (2015).

your luck must turn. Caution about our biases, combined with induction from past cases, suggests that if we believe our odds of success at T_n are as good as they were at T_1, we're typically kidding ourselves. In realistic cases otherwise similar to *Infinite Loop*, this explains the intuition that we should stop.

I am sceptical about attempts to explain away intuitions about one case by saying that we are simply mistaking it for another, different case. *Infinite Loop* stipulates that the chances are objective, and known. I see no problem holding this possibility in one's imagination. I can figure out my considered judgement on this case as readily as I can on the other hypothetical cases used to construct theories of the morality of self-defence and war. The objection insists that my intuitions about this case are the moral equivalent of a stubborn optical illusion. I don't buy it. What's more, the Discount View doesn't rest only on my considered judgement about *Infinite Loop*. It also draws support from the arguments of Section 4. I have a robust intuition that those who have already benefited from some sacrifices have a weaker claim to aid than those who have not, and this is enough to get the Discount View off the ground.

A second objection in this vein doesn't reach much beyond table-thumping, but is worth mentioning. Might one complain that it is absurd to suppose that, if I arrived on the scene of *Iterated Loop*, to find you standing there, haggard and conflicted, unable to decide whether to pull the lever, I should first ask you how many times you have pulled it? Isn't it

natural to view this as being, from my perspective, no different than if I were to come across the loop track just as the trolley starts to move for the first time?

I doubt that I will convince the incredulous. But there is nothing unusual about my proposal. Our present claims are often affected by the past—most obviously, when one has done something to make one responsible for the present situation. Entitlements grounded in historical acquisition, or in antecedent promises or contracts, might have a bearing on what one may do now. Only rarely can we take a decision problem at face value, without enquiring into its history. My proposal is in the same spirit as these familiar staples of deontological ethics.

So much for the Prospectivists' critique. From the other side, one might argue that the Discount View implies that the disvalue of causing a death changes, depending on whether it is ahead of or behind us.[30] One interesting feature of the Discount View, however, is that it is not committed to this idea. It says that your reasons *to save the five* in *Infinite Loop* are affected by the costs that you have already inflicted. That is consistent with insisting that the reasons against killing the one are invariant from one iteration to the next. Indeed, the Discount View allows us to capture the intuitively attractive thought—violated by the Prospective View—that we owe it to the

[30] Thanks to a referee for this objection.

victims of our failed attempts to save the five, to take their deaths into account at subsequent decision-points.

Some critics might reject the Discount View because they still don't buy its interpretation of trolley cases: despite my arguments to the contrary, they might insist that we don't have status-based reasons to save the five, only well-being-based reasons. I think this objection can take three different forms.

First, perhaps our reasons to save the five in *Iterated Loop* are wholly grounded in the intrinsic value of the well-being realised by saving their lives. Deontologists are unlikely to endorse this position. But if they do, then they should be Prospectivists in this case. I don't see why the intrinsic value of the beneficiaries' well-being should be affected by the moral costs incurred in failed attempts to save the five.

Second, perhaps we don't have any reason to promote intrinsically valuable well-being, and the only reasons at stake are respect-based. On that account, the Discount View still goes through—as I noted above, our status-based reasons (which on this view would be all of the relevant reasons) most likely diminish to an asymptote, rather than being wholly exhaustible.

Third, perhaps we only have well-being-based reasons to save the five, but those reasons are grounded both in the intrinsic value of the

beneficiaries' well-being, and in their moral status.[31] If that's right, then my arguments above can be interpreted as focusing on *the elements of our well-being-based reasons* that have to do with status on the one hand, and the intrinsic value of well-being on the other. The underlying ideas do not depend on my particular method of book-keeping. They depend only on the idea that part of the reason to save the five has to do with the intrinsic value of well-being, which cannot diminish in weight because of sunk costs, and part has to do with facts about status and respect, which can diminish in weight because of sunk costs.

The last two worries are less objections, than invitations to develop the details of the Discount View. As such, they must largely remain invitations: in the space remaining I cannot answer them satisfactorily.

The first asks how we determine precisely which costs count against our proportionality budget, and how they do so. Suppose, for example, that the first attempts to save the beneficiaries were incompetent, with very little prospect of success. Or suppose that they were intentionally abortive, with the malicious aim of using up the proportionality budget. Or, finally, suppose that the costs inflicted were extremely unlikely to occur—even wholly unpredictable—when you acted. How should we deal with each of these cases?

---

[31] Chappell (2015).

This objection raises tricky questions that any theory of proportionality must address. When considering the proportionality of a campaign to achieve some objective, we must ask which costs and benefits count towards that judgement. This is true even on a wholly synchronic version of proportionality—for example, when thinking about the proportionality of a military campaign, how should we factor in the innocent lives that will predictably be taken by incompetent or malevolent subordinates? How should we distinguish between costs inflicted by our side, and those inflicted by the other side? How do we account for totally unexpected costs (the 'unknown unknowns')? I cannot hope to answer these questions here, but I can gesture at some responses.

Most probably, costs inflicted by the incompetent or the malevolent should count less against the status-based reasons to aid the beneficiaries of your action than would costs inflicted in the sincere and competent attempt to save them. In effect, those costs should go on the 'moral ledger' of the incompetent and malevolent agents, rather than counting against the claims of their supposed beneficiaries. And as for unpredictable costs: if a given cost inflicted in the sincere and competent attempt to save the beneficiaries antecedently had a very low probability of coming about, then that should reduce the degree to which it discounts your status-based reasons to help those beneficiaries at a subsequent iteration. In other words, when weighing sunk costs against your proportionality budget, they must each be weighted for the antecedent probability that that particular outcome

would arise.

Consider a variation on *Iterated Loop*, in which there is a tiny probability at T_1 that instead of there being one innocent victim on LOOP, there are in fact 1,000. You pull the lever, and the trolley goes down LOOP, killing 1,000 people. I suggest that those deaths have considerably less effect on your status-based reasons at T_2 than they would have if their probability had been much higher. One plausible approach is simply to multiply the probability by the seriousness of the outcome; but a non-linear probability-weighting might be more plausible. Either way, the mere fact that something wholly unexpected and catastrophic happens need not entail that one is prohibited from continuing to try to save the beneficiaries.

Finally, when do the claims of the beneficiaries start to diminish? Are they limited to a particular choice situation or might they be determined by a longer time-frame? Suppose one of the prospective victims in *Iterated Loop* has been in the same situation on many different prior occasions, with many people's lives being sacrificed in part for his sake. Does this mean his claim at T_1 is weaker than that of others who haven't been the cause of significant costs being imposed on others?[32]

Obviously the intuitively correct response is that we should calculate the strength of each person's claim based only on the present choice situation. This raises the interesting question of how to delimit a particular

[32] Thanks to Christian Barry for helping me develop this objection.

choice situation, especially when one is engaged in a complex project like a war, which can be cut up into many different overlapping segments.

This points us towards a broader problem in normative ethics, for which everyone needs a solution. When we think about proportionality in general, what is the proper unit of analysis? In the ethics of self-defence we typically think that we should focus on a particular choice situation, rather than making overarching ethical judgements: the person who is responsible for this particular unjustified threat is potentially liable to be killed, even if on the whole he is a better person than the one whom he will otherwise kill. And one is normally only thought liable to be killed to avert a threat for which one is responsible oneself, rather than just to avert any comparably serious threat, imposed by anyone else (though this is more controversial). The Discount View raises no new problems here either.

6. Conclusion

Though the phenomenon it adverts to arises in many different areas of human agency, the moral sunk costs debate is in its infancy. Adherents of the Prospective View think it should stay there: this is a non-problem, since proportionality calculations must be strictly forward-looking. But while that move might appeal to a certain kind of causal consequentialist (of which there are few around these days), it makes little sense for the rest of

us. Obviously the past can affect what it is permissible to do now. And there is a limit to what we are owed by others, which must take into account the sacrifices already made on our behalf.

This does not mean, however, that we should allow moral sunk costs to dominate our reasoning. If you can save the lives of some at a small cost to others, then you are always permitted, and sometimes required, to do so. In these cases our well-being-based reasons, and plausibly also our status-based reasons, always have some weight. All the Discount View insists upon is that we cannot indefinitely continue to justify imposing the same expected costs on others, for the same probability of realising the same goal. At some point we must let those burdens fall where they will.

The Discount View is at least as credible as the Quota and Prospective Views: indeed, more so, since it can cater for all the intuitions that those views support, while avoiding the objections to both. What's more, I have given the Discount View a rationale in our underlying moral reasons (which has been done for neither of the alternative views). Our status-based reasons, grounded in the importance of showing equal respect to those of equal status, plausibly diminish in force as we rack up failed attempts to fulfil them. But we also have reasons grounded in the intrinsic value of the well-being of those whom we aim to save. And that intrinsic value is unaltered by the moral costs incurred when, as is sometimes inevitable, the world is uncooperative, and we can save some only at the cost of imposing risks on others.

Australian National University, Australia

References

Chappell, R. Y. (2015) 'Value Receptacles', *Noûs*, 49/2: 322-32.
Fabre, C. (2015) 'War Exit', *Ethics*, 125/3: 631-52.
Feinberg, J. (1978) 'Voluntary Euthanasia and the Inalienable Right to Life', *Philosophy & Public Affairs*, 7/2: 93-123.
Frowe, H. (2018) 'Lesser-Evil Justifications for Harming: Why We're Required to Turn the Trolley', *The Philosophical Quarterly*.
Hurka, T. (2005) 'Proportionality in the Morality of War', *Philosophy & Public Affairs*, 33/1: 34-66.
Jackson, F. and R. Pargetter (1986) 'Oughts, Options, and Actualism', *Philosophical Review*, 95/2: 233-55.
Kamm, F. M. (2001) 'Making War (and Its Continuation) Unjust', *European Journal of Philosophy*, 9/3: 328-43.
Kelly, T. (2004) 'Sunk Costs, Rationality, and Acting for the Sake of the Past', *Noûs*, 38/1: 60-85.
Lazar, S. (2016) 'Anton's Game: Deontological Decision Theory for an Iterated Decision Problem', *Utilitas*: 1-22.
McMahan, J. (2015) 'Proportionality and Time', *Ethics*, 125/3: 696-719.
McMahan, J. (2018) 'Proportionality and Necessity in *Jus in Bello*', in *The Oxford Handbook of Ethics of War,* Seth Lazar and Helen Frowe (ed.), New York: Oxford University Press: 418-39.
Moellendorf, D. (2015) 'Two Doctrines of Jus Ex Bello', *Ethics*, 125/3: 653-73.
Parfit, D. (2011) *On What Matters*, Oxford: Oxford University Press.
Rasmussen, K. (2012) 'Should the Probabilities Count?', *Philosophical Studies*, 159/2: 205-18.
Rodin, D. (2008) 'Two Emerging Issues of Jus Post Bellum: War Termination and the Liability of Soldiers for Crimes of Aggression', in *Jus Post Bellum: Towards a Law of Transition from Conflict to Peace,* Carsten Stahn and Jann K. Kleffner (ed.), The Hague: T.M.C. Asser Press: 53-76.
Rodin, D. (2015) 'The War Trap: Dilemmas of Jus Terminatio', *Ethics*, 125/3: 674-95.
Taurek, J. M. (1977) 'Should the Numbers Count?', *Philosophy and Public Affairs*, 6/4: 293-316.
Walden, K. (2014) 'The Aid That Leaves Something to Chance', *Ethics*, 124/2: 231-41.
Wasserman, D. (1996) 'Let Them Eat Chances: Probability and Distributive Justice', *Economics and Philosophy*, 12/1: 29-49.