The Philosophical Quarterly Vol. 0, No. 0 2020
ISSN 0031-8094 doi: 10.1093/pq/pqaa029

MORAL SUNK COSTS IN WAR AND SELF-DEFENCE

BY ELAD UZAN

The problem of moral sunk costs pervades decision-making with respect to war. In the terms of just war theory, it may seem that incurring a large moral cost results in permissiveness: if a just goal may be reached at a small cost beyond that which was deemed proportionate at the outset of war, how can it be reasonable to require cessation? On this view, moral costs already expended could have major implications for the ethics of conflict termination. Discussion of sunk costs in moral theorizing about war has settled into four camps: Quota, Prospect, Addition, and Discount. In this paper, I offer a mathematical model that articulates each of these views. The purpose of the mathematisation is threefold. First, to unify the sunk costs problem. Second, to show that these views differ in the nature of their justifications: some are justified qualitatively and others quantitatively. Third, to clarify the differential force of qualitative and quantitative critiques of these four views.

Keywords: moral sunk costs, self-defence, proportionality, iteration.

I. INTRODUCTION

Suppose it is determined that the taking of no more than 10,000 lives is justified in order to achieve the just goal of a given war, where those killed are not liable to be harmed. But suppose the war goes badly: 10,000 lives are lost while the goal remains unachieved, and it will take another 1,000 deaths to achieve it. Should the war stop? On the one hand, it seems obvious that it should: any further killing would make the total number of deaths disproportionate. On the other hand, the lives already lost are *sunk costs*: nothing can bring the dead back to life, so if achieving the just goal was worth 10,000 lives *ex ante*, surely it is worth 1,000 later. Both views seem compelling, so we face a dilemma. Should a war's sunk costs count towards the calculation of proportionality? And, if so, in what way?

Others have also asked these questions, resulting in several intriguing possible answers. Discussion of sunk costs in moral theorizing about war has settled into four camps, usefully categorized by Victor Tadros in his essay 'Past Killings and Proportionality in War' as the *Quota*, *Prospect*, *Addition*, and *Discount* views.[1] Suppose there is a war, starting at time T_1, in which country X plans to save 50,000 innocent persons from being killed by country Y's officials. Suppose as well that the standard 1:5 trolley-problem ratio obtains: if X prevents 50,000 deaths by killing 10,000 innocents, the killing is proportional.[2] But things do not go as planned, and there are early losses: X kills 10,000 innocents by time T_2, but the goal of preventing the 50,000 deaths has not yet been achieved. What should X do? The right choice depends on which of the camps one joins.

> *Quota:* If X's evidence warrants the belief that the total number of deaths that will be caused to save the 50,000, including early losses and those yet to come, will make the war as a whole disproportionate, then X ought not continue fighting at T_2.
>
> *Discount:* The fact that X has caused 10,000 deaths at T_2 counts against causing further deaths and might make further fighting widely disproportionate. But, in making the proportionality calculation at T_2, each death that has already occurred counts less than each prospective death.
>
> *Prospect:* The fact that X has caused 10,000 deaths in the effort to save the 50,000 does not count at all in the decision-making at T_2.
>
> *Addition:* The fact that X has caused 10,000 deaths counts in favour of continuing to fight at T_2. That so many have died already makes it proportionate to kill more people overall than was the case at T_1.[3]

In this paper, I offer a mathematical model that articulates each of these views. The purpose of the mathematisation is threefold. Firstly, to unify the sunk costs problem. Secondly, to show that these views differ in the nature of their justifications: some are justified qualitatively and others quantitatively. Thirdly, to clarify the differential force of qualitative and quantitative critiques of each of the four views.

The paper proceeds as follows. In Section II, I describe the close mathematical relationship among the four views. In Section III, I show that *Quota* and *Prospect* require only qualitative justification, while *Discount* and *Addition* require quantitative justification in addition to qualitative justification. In Section IV, I account for the risk of the *iteration problem*—that if one disregards sunk costs, past deaths can justify future deaths *ad infinitum*. In Sections II–IV, the only sunk costs I consider are deaths of those not liable to be harmed. Section V turns to those who are liable, evaluating whether and how lost combatants may be relevant to the sunk costs problem.

[1] Tadros (2018).
[2] But see Frowe (2018).
[3] Tadros (2018: 11–2).

In Section VI, I investigate the objection that non-*Quota* views are vitiated by the iteration problem and argue that it is mistaken. It is true that non-*Quota* views are susceptible to the iteration problem, but I show that the iteration problem is not intrinsic to non-*Quota* views. What *Quota* proponents oppose isn't non-*Quota* views as such but rather their misuse—it is not non-*Quota* views that justify endless fighting, but rather their misuse. I then show that in the typical case where *Prospect* or *Addition* is misused to justify continuing a war, (1) their correct use would be no more permissive than would correct use of *Quota*, and (2) that *Quota,* no less than *Prospect* or *Addition,* can be misused to justify war. Finally, addressing both the liable and non-liable cases, I tentatively show why there seems to be no reason except for the iteration—or misuse—problem to *a priori* exclude any of the four views. Section VII concludes the paper.

II. GENERALIZING THE PROBLEM OF MORAL SUNK COSTS IN WAR

Mathematically, *Quota, Discount, Prospect,* and *Addition* are on a continuum. Consider the following formula. CB represents the current proportionality budget—the number of lives remaining in our proportionality budget at T_2. OB represents the original proportionality budget at T_1. L is the number of lives already lost. x% tells us how much the past losses count—what their discount rate is. We get:

$$CB = OB - (L^{*}x\%)$$

In our example, OB = 10,000, and L = 10,000. So:

$$CB = 10{,}000 - (10{,}000^{*}x\%)$$

According to *Quota,* $x = 100$; all previous deaths count fully, so CB = 10,000 − (10,000 × 100%) = 10,000 − 10,000 = 0. Our future budget is 0, and any future death makes the war disproportionate.[4] In the case of *Discount,* $0 < x < 100$; the previous deaths are somewhat discounted. If they are discounted by 20% compared to future deaths, $x = 80\%$, and CB = 10,000 − (10,000 × 80%) = 10,000 − 8,000 = 2,000. At T_2, another 2,000 may be killed to save 50,000. In the case of *Prospect,* $x = 0$; all previous deaths are fully discounted, and CB = 10,000 − (10,000 × 0%) = 10,000 − 0 = 10,000. The previous deaths do not count at all; we still may kill up to 10,000 to save 50,000 at T_2.

[4] Here I am referring to evidence-based proportionality. Under all four systems, the war as a whole is fact-based disproportionate if more than 10,000 are killed to save 50,000.

Finally, in the case of *Addition,* $x < 0$. That is, those who are harmed count in favour of increasing the current budget. If, for example, $x = -20\%$, then our current budget is $\mathrm{CB} = 10{,}000 - (10{,}000 \times -20\%) = 10{,}000 - (-2{,}000) = 12{,}000$. We may kill up to 12,000 to save 50,000 at T_2.

The weight given to past losses in reducing future allowable losses runs from 100% (*Quota*) to a negative value (*Addition*). As we will see in the next section, placing the four views on this linear scale requires that we refine the justifications of each view.

III. QUALITATIVE AND QUANTITATIVE JUSTIFICATIONS

Quota and *Prospect* are extreme cases of *Discount*. *Quota* requires a justification for refusing to discount any sunk costs, and *Prospect* requires a justification for discounting all sunk costs. Neither requires further justification of the rate itself; such justification is implicit in the view. In other words, *Quota* and *Prospect* require only *qualitative* justification because both already imply a particular quantitative discount level.

There are many justifications for both views. Darrel Moellendorf argues that *Quota* is the only view that allows us to define fairly the 'cumulative costs of the war'.[5] Such cumulative costs should, according to Moellendorf, treat all those who died in the war as equal. It should not matter whether a death occurs at the beginning of the war or at its end; there is no discount on deaths that come first or last. Cecile Fabre adds that any discounting of earlier deaths would mean 'proportionality would lose most of its bite as a constraint against killing'.[6]

Supporters of *Prospect* argue that it is a fallacy to give any weight to sunk costs, in moral decision-making or any other kind. On this view, past losses should not matter in making forward-looking decisions. In David Rodin's words, since the war is already 'irredeemably disproportional', ignoring sunk costs leads one to choose the best way to mitigate the disproportionate harm done. Suppose at T_2 we must sacrifice another 1,000 to achieve the goal. We must do *something*—either continue fighting or end the war—but no matter what we do, the war will be disproportionate. The situation is a trap, Rodin says, something that is 'easy to get into but hard to get out of'.[7] But, he argues, it is morally better to lose 11,000 lives to achieve a just goal worthy of 10,000 lives than to lose those 10,000 lives for nothing. *Prospect* does not ignore the fact that causing the death of 11,000 to save 50,000 *is* disproportionate. *Prospect* merely denies that this disproportionality is a reason to exclude further fighting

[5] Moellendorf (2015).
[6] Fabre (2015).
[7] Rodin (2015).

a priori. Continuing the fight, though it be disproportionate, could make for the least-worst option.

The case of *Discount* differs from those of *Quota* and *Prospect* in that we must consider both qualitative and quantitative justifications. Qualitatively, *Discount* requires an explanation of why it is legitimate to have a discount level other than 0% (*Quota*) or 100% (*Prospect*). That is, we need to justify why is it legitimate to discount some but not all harm that has already occurred. Quantitatively, *Discount* requires an estimate, based on the amount of past harm, of the extent to which future harm should be discounted. Without a quantitative estimate, *Discount* is useless. The reasoning underlying that estimate demands justification.

In a recent paper, Seth Lazar gives a qualitative and quantitative defence of *Discount*. First he considers well-being-based reasons for fighting, which are justified by the goal of increasing the well-being of those we wish to help. He then considers four reasons grounded in equal moral status, which justify fighting because the persons we wish to help are valuable. He argues that status-based reasons become weaker as past costs mount, while well-being-based reasons do not deteriorate. This gives us a quantitative justification for *Discount:* since status-based reasons become weaker as costs mount, the discount rate of past lives is not 0%. But since well-being-based reasons do not become weaker as costs mount, the discount rate also is not 100%. The rate is somewhere in between. Since well-being-based reasons do not diminish at all, and in Lazar's view status-based reasons do not diminish very fast or very much, he concludes the overall discount rate is low.[8]

Like *Discount, Addition* requires both qualitative and quantitative justification. Qualitatively, we must explain what about the previous deaths makes them a source of justification for future killing. Quantitatively, we need to estimate how much added sacrifice is justified by past deaths. Consider Jeff McMahan's discussion of moral sunk costs in 'Proportionality and Time'.[9] Qualitatively, McMahan considers the deaths of those not liable to be harmed to be morally relevant to the justification of continued fighting because they died while trying to achieve the just cause.[10] Their well-being can no longer be changed, but their deaths can be 'partially redeemed by the subsequent achievement of the just cause'.[11] According to this *Redemption Thesis,* as McMahan calls it, 'the redemption of sacrifices made in the past can be an additional good that weighs in the assessment of the proportionality of continuing a war'.[12] Redemption can happen in various ways, in particular by making deaths more meaningful to those who survive: 'The dead soldier's parents, for example, seem entitled

[8] Lazar (2018).
[9] McMahan (2015).
[10] McMahan (2015: 698).
[11] McMahan (2015: 702).
[12] McMahan (2015: 703).

to a certain comfort if the just cause in pursuit of which he died is achieved, a comfort they would be denied if the war were lost'.[13] According to McMahan, redemption cannot count as a just goal on its own, independently justifying fighting once the original just goal has been achieved. But, as an additional good, it can count in the calculus of proportionality when deciding whether to continue fighting for the original just goal.

The question, then, is how redemption can justify the death of anybody who is still alive. Can the comfort of grieving parents possibly justify the risk that other parents will lose their own soldier sons and daughters? McMahan, aware of the moral force of this objection, argues that redemption must be narrow in scope. For him the redemption thesis can justify only the deaths of those liable to be harmed—for instance, enemy soldiers who participate in an unjust war. Redemption cannot justify the death of even one of one's own (just) soldiers, let alone the deaths of one's civilians, since none of them are liable to be harmed. Nor can redemption justify the death of non-liable enemy civilians, who have done little or nothing to support the unjust cause.[14]

Now, suppose that at T_2 10,000 non-liable persons were killed. Suppose we now estimate that another 10,000 just (non-liable) soldiers will have to die in battle to save the 50,000 and that victory will require destroying a crucial armament factory that was discovered after the war began. This will cause the death of 100 non-liable civilians living next to the factory. The total number of non-liable lives needed at T_2 to save 50,000 is 10,100. *Prospect* allows killing 10,000 more non-liable persons at T_2 in order to achieve the goal, but not 10,100. *Addition* would allow killing 10,100 more non-liable persons at T_2 since achieving the goal would partially redeem the death of the 20,100 total non-liable.

To give a less formal example, when Lincoln noted in the Gettysburg Address that 'from these honoured dead we take increased devotion to that cause for which they gave the last full measure of devotion', he knew this meant that more civilians, as well as more soldiers, might die to ensure those lost 'shall not have died in vain'.[15] This does not seem to make his view morally wrong. To claim *a priori* that the partial redemption of any number of non-liable persons who were already killed does not justify the death of even a single non-liable person seems too rigid. It is as rigid as claiming, under *Quota,* that not one more person may be killed to achieve the goal once *Quota* is reached.

[13] McMahan (2015: 711).

[14] McMahan (2015: 714). The non-liability of soldiers fighting for a just cause is a feature of revisionist just war theory, to which McMahan subscribes. Traditional just war theory sees all combatants as having an equal moral right to fight, while revisionist just war theory contends that combatants fighting for an unjust cause have no right to kill. Another difference is that while traditional theory invokes civilian immunity, revisionist just war theory does not, hence it is necessary, for purposes of this discussion, to specify that the enemy civilians are non-liable.

[15] Abraham Lincoln, Gettysburg Address, November 19, 1863.

But if the redemption thesis permits, qualitatively, *some* extra harm to those not liable to be harmed, *Addition* adherents need to figure out, quantitatively, *how much* it permits. The above example seems reasonable, since if 10,100 non-liable dead is proportional at T_2, $x\%$ is just -1%. To illustrate why there must be a limit, consider what *Addition* allows without one. A real-life example is the infamous Nazi order to kill 100 civilian hostages for every German settler killed in the occupied territories. This is obviously morally bad. It would also have been morally bad, though not quite *as* bad, for the Allies to claim that every Allied soldier's death justified killing an extra 100 German civilians in the cause of achieving the war's goal.

Let me clarify the similarities and differences between the two situations. The Allied soldier, being a just soldier, is not liable to be harmed. Neither is the German settler, being a civilian who is not contributing to the war effort.[16] In the settler's case, however, the 100 civilians are killed deliberately, and without any regard to achieving a just goal. They are killed merely in retaliation. In the Allies' case, the civilians are not targeted deliberately. They are only added to the proportionality budget of (non-liable) civilians who may be killed in order to achieve the just goal. This might be legitimate according to *Addition,* since achieving the just goal would now also achieve the secondary goal of redeeming the Allied soldier's death. But surely allowing another 100 civilians to be killed to redeem the death of one soldier is grossly disproportionate even if he fought for a just cause. It is *Addition* with $x\% = -9{,}900\%$. This would allow the death of an extra 990,000 non-liable people at T_2, if 10,000 just soldiers were killed and the goal was not yet achieved. Extreme *Discount* is reduced to either *Prospect* or *Quota,* but extreme *Addition* is extreme indeed. It seems that the extra harm that may be imposed on the non-liable to redeem those already harmed might not be zero but still must be rather small.

IV. SUNK COSTS AND THE ITERATION PROBLEM

The iteration problem is potentially an acute one because war is a process of sequential decision-making. As Henry Shue put it, 'One must both ask prior to going to war, would the evil to be prevented by military action in this case be worth engaging in a war overall, and ask throughout any war engaged in, would this particular military engagement make a sufficiently great contribution to

[16] Perhaps settlers in such territories are liable to be harmed despite not being soldiers, because settlement is forbidden. However, these people are not liable to be *killed*. What is more, the Germans gave orders to execute 100 hostages for every German *soldier* killed, and such soldiers were liable to be harmed, being unjust combatants. In any case, if the settler or soldier harmed were in fact liable, this would make the Nazis' orders even more unethical than in the example given.

potential victory to be worth the death and destruction likely to result?'.[17] If belligerents respond to sunk costs by re-evaluating upward the degree of harm they may justifiably inflict in order to achieve a morally worthy goal, then they can theoretically justify the infliction of harm without limit.

This carries significant implications for the ethics of conflict termination under a non-*Quota* model, in particular *Prospect*. As Lazar explains:

> Because [*Prospect*] 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. . . . 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.[18]

The answer to the iteration problem might seem to be *Quota*. If *Quota* is strictly followed, then we do not engage in sequential decision-making about proportionality. We decide only once, at the beginning of the war, how much harm is allowed. We never choose to stop the war. We merely execute an earlier decision to stop fighting regardless of the outcome. Like Odysseus, we have already tied ourselves to the mast. Even if one needs, at the end of the way, to sacrifice only one more soldier to save 50,000 innocents, one may not do so.

The way to avoid this dilemma—either allow fighting to continue forever or else demand rigid adherence to the original proportionality calculation—is to show that non-*Quota* views do not actually provoke an iteration problem. Let's begin with *Prospect*. On this view, we must always ignore past losses and continue to fight as long as the just goal is not yet achieved. We may therefore have to iterate: to sacrifice the expected number of victims over and over until we achieve the goal. The only way to avoid tiny chances of great harm is contingent pacifism, yet proponents of *Quota* are no more pacifists than are proponents of *Prospect*.

What about other non-*Quota* cases? It turns out that not only *Prospect* but all non-*Quota* views, including *Addition*, can be defended from the iteration problem using the mathematical model. To illustrate this, I borrow the mechanics of Lazar's *Iterated Loop* case:

> 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, p-1, 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

[17] Shue (2005: 748).
[18] Lazar (2018: 847).

be replaced, leaving you with the same decision at T2 as you faced at T1, with just the same odds; the same holds for T3–Tn.[19]

Now suppose n is unlimited; there is an infinite supply of potential innocent victims to place on the LOOP track.[20] Let us approach the problem using each of three of our views: *Quota*, *Prospect*, and *Addition* with $x\% = -50\%$.[21] Assume, further, the trolley problem's 1:5 ratio between deaths caused and deaths prevented. Express costs and gains in units of deaths caused: if we kill one person, the cost is 1 life; if we prevent the death of 5 people, the gain is 1 life. If $p = 0.5$, *Quota*, *Prospect*, and *Addition* are indifferent as to whether we should pull the lever. If we pull the lever at T_1, we have a 50% chance of preventing 5 deaths (expected benefit: $50\% \times 1 = 0.5$) and a 50% chance of causing one death (expected cost: $50\% \times (-1) = -0.5$), for a total expect utility of 0. There have been no previous deaths, so *Quota* and *Addition* make the same recommendation as *Prospect*.

Suppose now the trolley goes on the LOOP track at T_1. It follows that we are now at T_2, when we already caused one death. The recommendations of *Prospect, Quota,* and *Addition* now diverge. At T_2, the cost already incurred is 1, so *Quota* tells us to stop. This cost is already equal to the maximal gain that could be achieved at T_1. *Prospect* tells us we may, but need not, pull the lever at T_2, or at any T_n for $n > 2$. The expected cost is still 0.5, and the expected gain 0.5. Sunk costs do not matter. No matter how many times the trolley goes on the LOOP track, at T_n it is permissible either to pull the lever or not.

Addition tells us we *must* pull the lever at T_2. Now the expected gain of diverting the trolley to STOP is not just $0.5 \times 1 = 0.5$ (a 0.5 chance of preventing the death of 5 people), but $0.5 \times (1 + 0.5 \times 1) = 0.75$. The reason is that if we succeed in diverting the trolley, we gain not only the benefit of preventing the death of five people but also redeem the death of the person killed in the first iteration. The expected cost of pulling the lever and making the trolley go through LOOP again is still the same, however: $0.5 \times (-1) = -0.5$. What is more, for $n > 2$, every further LOOP victim will only make the expected gain from pulling the lever even higher than at T_2.

Thus, for any $n \geq 2$, *Prospect* allows, and *Addition* requires, pulling the lever. This seems to justify the fears about iteration. We may, or even must, pull the lever an indefinite number of times, killing a potentially unlimited number of people. But this ignores a crucial question: What is the *chance* that puling the lever will make the trolley go through LOOP two, three, or more times? While

[19] Lazar (2018: 845ff).

[20] Lazar (2018: 847).

[21] This number is chosen arbitrarily, but the argument is the same no matter what $x\% > -100\%$ is chosen for *Addition*. This lower bound is not a significant restriction on $x\%$, since any reasonable $x\%$ for *Addition* is likely to be relatively low.

the cost of repeatedly pulling the lever and diverting the trolley to the LOOP track rises linearly, the chance of this happening declines exponentially. There is a one-in-two chance of the trolley going through LOOP once, but only a one-in-four chance of it going through LOOP twice, one-in-eight of it doing so three times, and so on.

With this in mind, we can calculate the expected costs and benefits from repeated application of *Prospect* and *Addition*. For each *n*, we need to consider two things. Firstly, what is the chance the trolley reaches STOP at the *n*th iteration (P_n)? Secondly, what would the overall benefit (B_n) minus the overall cost (C_n) be if that happens? The infinite sum of these quantities, $\sum_{n=1}^{\infty} \mathrm{P}n * (\mathrm{B}n - \mathrm{C}n)$, tells us the expected utility of repeatedly pulling the lever until STOP is reached.

In the case of both *Prospect* and of *Addition*, the chance of the trolley reaching the STOP track at the *n*th pull of the lever, P_n, is 0.5^n. It can happen only if the trolley goes *n*-1 times to the LOOP track, then one final, *n*th, time, moving to the STOP track. The cost of *Addition* and *Prospect* is the same: $n - 1$ people killed—one for each time the trolley has gone through the LOOP track. So in both cases $C_n = (n - 1)$. The difference is in the benefits. In the case of *Prospect,* the benefits are always the same: $B_n = 1$ (preventing the death of five people in the last iteration, when the trolley moved to STOP). In the case of *Addition,* the benefits are higher. We not only save five people (utility: 1) but gain $0.5 * (n - 1)$ in additional benefits, since we redeem the death of the previous n-1 victims of LOOP. The expected utility in *Prospect* is

$$\textit{Prospect}: \quad \sum_{n=1}^{\infty} [\mathrm{P}n * (\mathrm{B}n - \mathrm{C}n) = \sum_{n=1}^{\infty} (0.5)^{\mathrm{n}} * (1 - (\mathrm{n} - 1)) = \sum_{n=1}^{\infty} (0.5)^{\mathrm{n}} * (2 - \mathrm{n}) = 0$$

In other words, the mere fact that there *might* be a great harm does not make *Prospect*'s utility negative. What is more, even if we look only at the expected harm, it is $(-1) \times (\sum_{n=1}^{\infty} (0.5)\hat{n}) = -1$. There is a chance of many people being killed by repeated failures, but the expected overall harm is only one death. The reason is that the chance of losing many lives quickly becomes very small. There is a one-in-two chance of killing one person but only a one-in-four of killing two, a one-in-eight chance of killing three, and so on. The expected overall benefit is $1 = (\sum_{n=1}^{\infty} 0.5\hat{n})$. There is a one-in-two chance of saving five people (moral benefit = 1) on the first pull of the lever. There is a one-in-four chance of saving five people on the second pull of the lever, a one-in-eight chance of saving five people on the third pull of the lever, and so on. Thus the expected number of lives saved is five, which means the moral gain is one. What's more, if one keeps pulling the lever, the trolley will eventually move to STOP and save five lives. So, on average, *Prospect*'s expected moral utility of *Iterated Loop* is 0: expected gain of 1, expected loss of −1.

Under *Addition,* the expected utility is:

$$\textit{Addition}: \quad \sum_{n=1}^{\infty} \mathrm{P}n * (\mathrm{B}n - \mathrm{C}n) = \sum_{n=1}^{\infty} (0.5)^{\mathrm{n}} * \ ((1 + (0.5)(\mathrm{n}-1) - (\mathrm{n}-1))$$
$$= \sum_{n=1}^{\infty} (0.5)^{\mathrm{n}} * \ (1.5 - 0.5\mathrm{n}) = \ 0.5$$

Addition has a higher expected utility than *Prospect.* If we succeed in diverting the trolley, not only do we save five innocent people, but we also partially redeem the lives of those killed in the previous, failed attempts. But note that the entire contribution of all the expected additional benefits of redemption is the third series in the formula, $\sum_{n=1}^{\infty} 0.5(\hat{n} + 1)$, that is, only 0.5 lives. This is because *Addition*'s expected harm is only one life—the same as *Prospect*'s, and for the same reason. This in turn also limits the expected redemption to only 50% of this expected harm: 0.5 lives. *Addition*'s utility is higher than *Prospect*'s since it is possible that *Addition* will redeem the lives of those lost in failed attempts to divert the trolley. But to do so, the trolley would have to first cause a lot of harm by failing to stop many times, and we have seen that the chances of that are very small. There is a one-in-two chance of having no lives to redeem (the trolley moves to STOP on the first try), a one-in-four chance of redeeming one death (the trolley moves to STOP on the second try), a one-in-eight chance of having two lives to redeem, and so on. *Addition* only justifies sacrificing an extra half of one life, since the cost of stopping at $\mathrm{T_n}$ grows linearly, but the chance of paying such a cost and not reaching STOP declines exponentially.

So in cases of both *Addition* and *Prospect,* the chance of paying a very high cost is negligible. To give one example, the chance of having 100 or more deaths is 1 in 2^{100} to about 1 in 10^{30}, or one chance in a hundred thousand billion billion billions. So the quantitative difference between *Quota, Discount, Prospect,* and *Addition* is not as great as one might think at first sight. The fear that *Addition* would allow unlimited continuation to achieve a just goal is as exaggerated as the fear that *Prospect* would. Thus the quantitative argument against non-*Quota* views—that denying *Quota* would empty proportionality of its moral force—is a chimera. The fear of an infinite series of continued attempts to achieve the just goal is a fear only of logical possibilities—Cartesian doubts, to paraphrase C. S. Peirce.[22]

V. SUNK COSTS AND THE NUMBER OF COMBATANTS

So far I have dealt only with harm inflicted on those not liable to be harmed. In what way, if at all, is the problem of sunk costs relevant to harm inflicted on those liable? In order to answer this question, we need to first think more carefully about who is liable to be harmed in war and to what degree.

[22] Peirce (1877).

In this respect, it is useful to consider McMahan's distinction between narrow and wide proportionality. Narrow proportionality applies to those liable to be harmed due to their acts of aggression, in particular unjust enemy combatants.[23] This marks a refinement of traditional just war theory, in which all soldiers are liable to be harmed—any number of soldiers may be killed, without proportionality consequences.[24] Under the regime of narrow proportionality, we have to consider the lives of soldiers, just and unjust, in proportionality considerations. Wide proportionality applies only to those not liable to be harmed: innocent bystanders and one's own just combatants. The harm they experience in war can only be justified as a lesser evil.

Narrow proportionality applies individually to each liable person, by 'pairwise comparison'.[25] If it is narrowly proportional to harm one soldier who is unjustly attacking you, it is narrowly proportional to harm each of the 1,000 soldiers unjustly attacking you. The fact that another 999 soldiers are also unjustly attacking you makes no moral difference with respect to what is permissible in defending yourself against the 1,000th soldier.[26]

Recent work in the ethics of war suggests that narrow proportionality's individualistic notion of liability leads to a puzzle.[27] Consider two types of liable killers: *culpable* ones and *responsible* ones. Culpable killers are fully morally responsible for the harm they intend. One example might be a Mafia hitman. Responsible killers are those who bear only partial moral responsibility for the harm they inflict. These include, say, a driver of a runaway vehicle that strikes and kills a pedestrian.[28] Both killers are liable to be harmed in order to prevent harm to oneself or others. However, it seems intuitively obvious that the harm aggregates only in the case of responsible killers and not of culpable ones. If you may kill one mafia hitman, you may kill 1,000. You may kill one driver to save yourself, but you may not kill 1,000.[29]

The puzzle arises from the realization that narrow proportionality makes no distinction between the two cases. If it is narrowly disproportionate to harm

[23] McMahan (2009: 183).

[24] The *locus classicus* account of the moral equality of combatants is in Walzer (1977: 36–7).

[25] McMahan (2017: 24).

[26] As an aside, I note that combatants rarely engage in such pairwise comparisons and instead consider enemy soldiers as a group. There are rare cases, however, in which the number of unjust soldiers does reduce the liability of each individual soldier. For instance, if 500 reluctant conscripts are guarding ten prisoners, each guard's degree of responsibility is too small to render any one of them liable to be harmed. For a formulation of this case, see McMahan (2009: 23).

[27] McMahan (2017: 5–6). Rodin (2017).

[28] This distinction is elaborated in McMahan (2017: 5–6).

[29] It is conceivable that all killing, even of highly culpable killers in necessary self-defence, does some unjustified moral harm. Even the most culpable are mere mortals, whose lives retain some moral worth, no matter how blameworthy their actions, or how necessary it is to kill. If so, the harm done by necessarily killing highly culpable killers aggregates. It does so, however, at a much slower rate than that done to merely responsible killers. Thanks to an anonymous referee for pushing me on this point.

a particular driver among 1,000—because doing so contributes little to one's chances of averting an unjust threat—then it is narrowly disproportionate to harm a particular Mafia hitman out of 1,000 for the same reason. Conversely, if it is narrowly proportional to harm one driver, it is narrowly proportional to harm them all.

Both horns of this dilemma are unacceptable. Call this the *numbers problem*. The numbers problem is especially important since, in most wars, most liable soldiers are merely responsible, not culpable. They are not wholly responsible for the harm the war inflicts. They may be conscripts, or fully believe the harm they do is just. In order to evaluate the extent to which harm to liable persons should be considered in sunk-cost calculations, we need a measure of proportionality that takes into consideration the difference between culpability and responsibility.

Several authors—among them Bazargan (2014); Rodin (2017); and Tomlin (2020)—have developed solutions to versions of the numbers problem. Each solution has its merits and shortcomings. I myself am not inclined fully to support Bazargan's, Rodin's, or Tomlin's solutions, for reasons independent of the problem of moral sunk costs.[30]

McMahan offers his own solution, a measure of proportionality that accounts for the numbers problem. He calls this *proportionality in the aggregate*. This measure of proportionality recognizes the killing of responsible soldiers as potentially disproportionate.[31] Proportionality in the aggregate compares the collective harm all killers will inflict to the total relevant moral benefits that would accumulate if all of them were stopped.[32] The key motivation for this form of proportionality is to minimize the amount of unjust harm done in circumstances when harm is unavoidable.[33] Consider 1,000 drivers who lose control of their cars. Each driver knowingly took a risk of harming others when he got behind the wheel. If the driver hits a pedestrian, he is more liable for any harm than the pedestrian is, although the pedestrian also took some risk upon himself when going out in the street. Say, then, that the driver is liable for 90% of the risk and the pedestrian from the remaining 10%.

So if 1,000 responsible killers (the drivers) were killed in order to save one pedestrian, the result would be $10\% \times 1{,}000 = 100$ innocent lives lost. Clearly these drivers should not all be killed in order to save one pedestrian: it would be disproportionate in the aggregate to lose 100 innocent lives to achieve the just cause of saving 0.9 ($90\% \times 1$) innocent lives. Now, consider the case of culpable killers. Imagine 1000 Mafia hitmen, one after another, trying to kill an

[30] Owing to space constraints, I cannot dissect and respond to their arguments here, though intend to in future work.

[31] McMahan (2017: 24).

[32] McMahan (2017: 25).

[33] McMahan (2011a).

innocent victim. Here, killing all of the hitmen leaves 'no residual injustice'.[34] Each killer is 100% liable, so killing any or all of them results in 0% harm to innocents—no unjust harm.

Supposing proportionality in the aggregate is the correct measure to determine how much harm may be done to liable persons in war, we now must determine whether our moral goals are worth the moral costs. Suppose it is proportional in the aggregate to kill 100 enemy soldiers to achieve a just cause. Suppose as well that, having killed the 100 enemy soldiers, you find your cause has not been advanced at all. How should your moral sunk costs affect what you ought to do next?

Consider the following case:

> *Simplified Invasion:* At time T_1, Attacker invades and occupies Defender's territory. The invasion is unprovoked and has no justification, thus all of Attacker's soldiers fighting in the invasions are liable to be harmed. No civilians are harmed: the invaded territory is inhabited only by soldiers.

Simplified Invasion is simplified in two ways. Firstly, most real wars involve civilians, even when they take place in remote places or at sea, where civilian cargo ships might be at risk. For our purposes, though, it is important that only liable combatants are involved. Second, there are no ambiguities with respect to justification and provocation. The purpose of this simplification is to position 100% just combatants (not liable) against 100% unjust combatants (liable but not necessarily culpable).

Suppose that it would be proportional in the aggregate to kill 10,000 of Attacker's soldiers in self-defence. Suppose further that at a later T_2, 10,000 of Attacker's soldiers have been killed, yet Attacker fights on. In analogy with the wide proportionality case, and using Tadros's formulation, there are four possibilities:

> *Quota:* If Defender's evidence warrants the belief that defending against the invasion will entail killing so many of Attacker's soldiers—those already dead and those to come—that the war as a whole will become disproportionate in the aggregate, then Defender ought not continue fighting at T_2.

> *Discount:* Killing 10,000 of Attacker's soldiers in self-defence by T_2 counts toward further fighting becoming disproportionate in the aggregate. But, in making the proportionality calculation, each death that has already occurred counts less than each prospective death.

> *Prospect:* Killing 10,000 of Attacker's soldiers in self-defence by T_2 doesn't count at all in the decision-making at T_2. If it was proportionate in the aggregate to kill 10,000 of Defender's soldiers at T_1, it is still proportionate in the aggregate to kill (another) 10,000 of Attackers' soldiers at T_2.

[34] McMahan (2011b), summarizing his conclusion in McMahan (2011a).

Addition: Killing 10,000 of Attacker's soldiers in self-defence by T_2 counts in favour of continuing to fight at T_2. That so many liable soldiers have died already can make it proportionate in the aggregate to kill even more of Attacker's liable soldiers than the 10,000 allowed at T_1.[35]

What is the potential for iteration? Suppose that for each wave of Attacker's soldiers killed, there is a 50% chance that Defender will defeat the invasion and win the war. This situation is mathematically analogous to that of LOOP and STOP. The expected number of Attackers' soldiers killed is 10,000 in the case of *Quota*; 15,000 in the case of *Prospect*; somewhat higher in the case of *Addition* (17,500 if $x = -50\%$); and between 10,000 and 15,000 in the case of *Discount*, depending on the discount level. While the costs in harmed liable soldiers rise linearly, the chance of such costs being incurred decreases exponentially. There is a one-in-two chance of 10,000 of Attacker's soldiers dying, a one-in-four chance of 20,000 deaths, a one-in-eight chance of 30,000, and so on. Therefore the expected number of deaths is only 15,000.

It is true that future battles are almost never simple iterations of past ones.[36] Both sides learn from past actions, their tactics change, their resources get depleted (or sometimes grow), and the chances of success change. Yet, so long as Defender's chance of winning and stopping the war in each step is reasonably high, iteration will mean that the chance of Defender having to fight more than a few battles quickly becomes very small.

VI. THE RISK OF MISUSE

In my view, the defensible argument against all non-*Quota* views—whether we are concerned with traditional measures of proportionality or with proportionality in the aggregate—is not that they are subject to iteration. The defensible argument is that decision-makers will respond to sunk costs by making unethical choices. In other words, the problem is that non-*Quota* views are subject to abuse. The preference that Fabre and others express for *Quota* is a preference for the option perceived as least vulnerable to abuse.[37]

We can be fairly certain that the problem with non-*Quota* options is external to them—i.e., one of misuse—because in nearly all realistic cases, each of the options, properly applied, would yield similar results. Return for a moment

[35] This conclusion is not as counterintuitive as it seems. It is better, morally speaking, that 10,000 soldiers die for nothing than to achieve an unjust goal, as a death for that goal would have negative moral value. *Addition* would therefore argue that if 10,000 soldiers die in an unjust cause, further sacrifice may be warranted to ensure that the unjust goal is not achieved, which would partially redeem the soldiers' deaths. This is analogous to *Addition*'s claim that additional sacrifice can be worthwhile if it ensures that just soldiers have not died in vain.

[36] Thanks to an anonymous referee for this objection.

[37] Fabre (2015: 637).

to the LOOP scenario. Suppose there is in fact little chance that further sacrifice will engender success—for instance, if the real chance of diverting the trolley to the STOP track is 10%. The expected gain of *Prospect* in this case is 1 because eventually five deaths will be prevented. The expected cost is $\sum_{n=1}^{\infty} (0.9)^{\mathrm{n}} * (-1) = -9$, for an expected utility of -8 lives. *Addition*'s utility would be $\sum_{n=1}^{\infty} (0.9)^{\mathrm{n}-1} * (0.1) * (1.5 - n) = -4.5$. As expected, higher than *Prospect*'s, but still very low. What is more, as the actual chance of success gets closer to 0, the expected utility becomes exponentially smaller: if there is a 1% chance of success, then the expected utilities of *Addition* and *Prospect* are -49.5 and -98, respectively. So iteration doesn't appear to be a problem. The low chance of future success as compared to future harm means that *Prospect, Addition,* and *Discount*—like *Quota*—would all recommend against diverting the trolley in the first place and against continuing to do so for $n \geq 2$.

The reason non-*Quota* views seem so much more permissive is that combatants often assume they have a higher chance of success than they really do. This is the case even though repeated failure to achieve a morally worthy goal—the condition of iteration, in other words—is evidence that the war's chances of success are low.[38] For this reason, decision-makers should consider the need to iterate as a strong evidence that the war is not proportional. As McMahan notes, 'If a government was repeatedly mistaken in its assessments of proportionality, the explanation would almost certainly be that it was incompetent or biased in making its predictions rather than that it was the victim of a statistically improbable series of epistemically justified judgments that all unluckily turned out to be mistaken'.[39]

But the misuse of *Prospect, Addition,* or *Discount* by grossly overestimating the probability of success in each iteration is no argument against their legitimate use. Also, it does not seem the misuse of non-*Quota* views is necessarily more prevalent than of *Quota*. For example, if one declares that the cause of a war is so important that almost any sacrifice would be justified, then *Quota* presents no obstacle to endless war. One could find many other cases to illustrate the point, not least because the institutions tasked with war-making decisions are typically headed by individuals—politicians and generals—whose career interests benefit from war.

So while it is possible that *Addition*, *Prospect*, and *Discount* are easier to abuse than *Quota*, the claim is debatable from an empirical standpoint. And, even if it is true, it does not follow that implementation of non-*Quota* views is illegitimate. Non-*Quota* views are not *a priori* excluded, either by the iteration problem or the risk of abuse. Rather than dispense with *Addition, Discount,* and *Prospect,* we should think about when they are appropriate.

[38] See, e.g., Lazar (2018: 857).
[39] McMahan (2015: 707).

Flexibility in thinking about sunk costs is valuable, especially when considering the case of harm to those liable. In such cases, *Quota* can produce unreasonable outcomes. In *Simplified Invasion*, Defender could, under *Quota*, 'buy' the territory occupied by Attacker by allowing 10,000 of its soldiers to be killed. If *Quota* is the only way to avoid an unlimited continuation of war, we may have to accept this repugnant conclusion; at the very least, we would be facing a serious dilemma. But, in fact, there are other options. Under *Prospect*, for example, Defender could still 'buy' the territory occupied by Attacker for the expected price of 15,000 soldiers, though it would have to take into consideration the possibility of higher costs. From the standpoint of minimizing bloodshed, this is not a better outcome than expected from *Quota*. But nor does *Prospect* result in the feared limitless war.

The risk of misuse of non-*Quota* views is mitigated by one more factor: retrospective duties.[40] Suppose that the original proportionality budget of 10,000 lives is exhausted. Before decision-makers can seriously consider what to do next, they need to fulfil their retrospective duties. These include determining why their original estimate was wrong, including investigating (*inter alia*) if they were negligent or, worse, intentional in underestimating the moral costs of achieving the goal. Such an analysis is highly relevant to re-estimating how many future casualties can be expected before the goal will be achieved. Only after making this analysis in good faith do the decision-makers have enough information to determine what their prospective duty is—that is, whether to continue fighting.

This retrospective analysis should have mitigating effects because the very fact that the goal was not achieved is *prima facie* evidence that achieving the goal is costlier that originally estimated. Therefore the number of future casualties now needed to achieve the goal should *prima facie* be adjusted upwards, too. This makes it less likely that it is (prospectively) proportionate to continue the fighting, whether the decision-makers apply *Addition*, *Prospect*, or *Discount*. (*Quota* always requires stopping once the original estimate is reached.)

To use a more specific example, suppose that after 10,000 casualties—the proportionality budget—the goal has only been 10% achieved. Let us continue to assume, as I have thus far, that the source of failure is mere bad luck. Decision-makers therefore need revise their estimate; the cost of achieving the goal is still 10,000. Since 90% remains to be achieved, another 9,000 casualties are needed. *Quota* (always) and *Discount* (usually) will forbid continuing, while *Prospect* and *Addition* permit it. Even so, as we have seen, the probability that there will be many more casualties is very low. It is hard to repeat such bad luck.

[40] I thank an anonymous referee for encouraging me to explore this possibility.

But maybe bad luck is not to blame. If the goal is indeed only 10% achieved after 10,000 casualties, this is strong evidence achieving the original goal requires many more casualties than originally estimated—that poor planning, not just misfortune, is at work. At this point, decision-makers could naïvely assume that a 10% achievement of the goal after 10,000 casualties implies 100,000 casualties to achieve 100% of the goal. This would mean 90,000 more casualties are needed. Under these circumstances *Quota*, *Prospect*, *Discount*, and *Addition* (unless very extreme) all demand an end to the war. Of course, actual revision of the proportionality budget is not likely to be so naïve. But, one way or another, failure to achieve the goal at the expected cost is *prima facie* evidence that the overall cost of achieving the goal should be revised upward, perhaps drastically. This is a curb on violence built into non-*Quota* views—as long as decision-makers behave ethically.

VII. CONCLUSION

Prospect, *Addition*, and *Discount* all require justification of harm beyond that allowed by *Quota*. *Discount* requires justification of exactly why and to what degree past deaths are discounted when one evaluates future allowable deaths. *Addition* requires justification of why and to what degree previous deaths can increase the number of future allowable deaths.

Using a simple mathematical model, I find that countenancing more harm than *Quota* allows will not cause runaway escalation in the degree of morally permissible harm during war. If the iteration argument fails against non-*Quota* views, then the viability of those views is strengthened. This has important implications for the consideration of moral sunk costs in war because, under *Quota*, such consideration is impermissible. Overcoming misdirected fear of non-*Quota* views creates space for other ethical possibilities. It may be that, as a matter of facts, sunk costs can ethically sanction continuation of a conflict beyond the point where the original proportionality budget is exhausted. But this need not mean that just war theory loses its capacity to constrain war-making. The constraint arises from sober, good-faith evaluation of the facts, a process that may be wanting under any ethical system.[i]

[i] For endless patience, support, guidance, and inspiration, I am grateful to Eyal Benvenisti and Moshe Halbertal. I conducted research for this article at Oxford's Faculty of Philosophy. My father, Jacob, passed away during my time there. Jeff McMahan was extraordinarily generous with his time, comments, and emotional support. For helpful and multiple discussions, I am particularly indebted to Gabriella Blum, Yitzhak Benbaji, and Arthur Ripstein. Finally, I am grateful to *The Philosophical Quarterly*'s editor and two anonymous referees for their thoughtful suggestions and criticisms.

ACKNOWLEDGEMENT

I wish to thank The Zvi Meitar Center for Advanced Legal Studies and the ERC-funded the *GlobalTrust – Sovereigns as Trustees of Humanity* research project—both at Tel Aviv University, and to the President of Israel Doctoral Fellowship for Scientific Excellence and Innovation, for their generous support. I conducted research for this article at Harvard Law School Program on International Law and Armed Conflict (HLS PILAC), and the Faculty of Philosophy at the University of Oxford. My sincere thanks to each of these institutions.

REFERENCES

Bazargan, S. (2014) 'Killing Minimally Responsible Threats', *Ethics*, 125: 114–36.

Fabre, C. (2015) 'War Exit', *Ethics*, 125: 631–52.

Frowe, H. (2018) 'Lesser-Evil Justifications for Harming: Why We're Required to Turn the Trolley', *The Philosophical Quarterly*, 68: 460–80.

Lazar, S. (2018) 'Moral Sunk Costs', *The Philosophical Quarterly*, 68: 841–61.

McMahan, J. (2009) *Killing in War*. Oxford: OUP.

——— (2011a) 'Who is Morally Liable to be Killed in War', *Analysis*, 71: 544–59.

——— (2011b) 'Duty, Obedience, Desert, and Proportionality in War: A Response', *Ethics*, 122: 135–67.

——— (2015) 'Proportionality and Time', *Ethics*, 125: 696–719.

——— (2017) 'Liability, Proportionality, and the Number of Aggressors', in S. Bazargan and S. Rickless (eds.) *The Ethics of War: Essays*, 3–27. Oxford: OUP.

Moellendorf, D. (2015) 'Two Doctrines of Jus ex Bello', *Ethics*, 125: 653–73.

Peirce, C.S. (1877) 'The Fixation of Belief', *Popular Science Monthly*, 12: 1–15.

Rodin, D. (2015) 'The War Trap: Dilemmas of *Jus Terminatio*', *Ethics*, 125: 674–95.

——— (2017) 'The Lesser Evil Obligation', in S. Bazargan and S. Rickless (eds) *The Ethics of War: Essays*, 28–45. Oxford: OUP.

Shue, H. (2005) 'War', in Hugh LaFollette (ed.) *The Oxford Handbook of Practical Ethics*, 734–58. Oxford: OUP.

Tadros, V. (2018) 'Past Killing and Proportionality in War', *Philosophy & Public Affairs*, 46: 9–35.

Tomlin, P. (2020) 'Distributive Justice for Aggressors', *Law and Philosophy*, (forthcoming): 1–29.

Walzer, M. (1977) *Just and Unjust Wars*, Fourth edition. New York: Basic Books.

Tel Aviv University, Israel