

# Duty, Obedience, Desert, and Proportionality in War: A Response\*

*Jeff McMahan*

---

This essay responds to four commentaries on my recently published book, *Killing in War*. It defends the view that soldiers ought to disobey an order to fight in a war that lacks a just cause, argues against the contractarian approach to the morality of war, develops an explanation of how the number of people who are harmed by defensive action can affect whether that action is proportionate in the “narrow” sense, and seeks to rebut the suggestion that an attacker’s desert may be relevant to the justification for harming him in self-defense.

---

I have sought to defend an account of the just war that is revisionist in a variety of ways. The commentators in this symposium represent a range of views that differ to varying degrees from my own. Although he develops an argument that ostensibly defends traditional just war theory, Cheyney Ryan ultimately rejects just war in favor of pacifism. Yitzhak Benbaji, by contrast, offers a robust defense of all the major elements of the traditional theory. David Rodin is an ally in the defense of a revisionist account, though his view seeks to retain certain elements of the traditional theory that I reject. Although John Gardner and François Tanguay-Renaud do not explicitly discuss war, they defend an account of individual self- and other-defense that, if applied to the conduct of war, might take them rather close to Ryan’s pacifism. I am enormously grateful to all five for their generous remarks and perceptive, unpolemical, and constructive criticisms and proposals. I will do my best to respond to some of their objections and to advance the discussion by building on their suggestions.

\* I am grateful to Seth Lazar and the Oxford Institute for Ethics, Law, and Armed Conflict for organizing and administering the conference at which the essays in this symposium were first presented. I owe further thanks to Lazar for insightful comments on the notes on which this essay is based, to Henry Richardson for comments on the penultimate draft, and to John Gardner and Cheyney Ryan for helpful correspondence.

*Ethics* 122 (October 2011): 135–167

© 2011 by The University of Chicago. All rights reserved. 0014-1704/2011/12201-0005\$10.00

## I. RYAN: OBEDIENCE VERSUS CONSCIENTIOUS REFUSAL

I have argued that it is impermissible to fight in an unjust war, by which I will here mean a war that lacks a just cause.<sup>1</sup> Ryan develops an argument—the *argument to democratic duty*—that concludes that soldiers in democratic states have a duty to fight in unjust wars. If this argument is successful, it would seem to undermine not only my view but Ryan’s pacifism as well. But perhaps Ryan assumes that it is only *if* we have armies that their members will have the duty to fight. Since he thinks we ought not to have armies, his defense of the argument is merely conditional.

One element of Ryan’s defense of the argument is an appeal to an analogy between wars and battles. Even in a just war, there may be certain battles that are unjust. Similarly, a war that is unjust may be a component of a larger enterprise that is just. Ryan assumes that just war theory does not require just combatants to distinguish between just and unjust battles and to fight only in the former. But if it is permissible for a combatant to fight in an unjust battle provided that it is a component of a war that is just, it should also be permissible for a combatant to fight in an unjust war if it is a component of an enterprise that is just.

The idea that the Iraq and Afghanistan wars are “fronts” in a larger “War on Terror,” or that the Korean and Vietnam wars were phases of a protracted war against communism, is incompatible with accepted criteria for the individuation of wars, including the criteria found in international law. But even if we grant this idea, there are reasons why combatants can reasonably be expected to be more selective about the wars in which they participate than they can be about the battles in which they fight. This is in part because it is more reasonable to assume that the purpose of a battle is given by the purpose of the war of which it is a part than it is to assume that the purported just cause of a war can be inferred from the purpose of some more encompassing activity of which the war is a component. A war is necessarily against specific adversaries, who must, if the war is just, have made themselves liable to be warred against. Once a just war has begun, there is therefore a presumption that any engagement with the adversary’s military forces is also just by virtue of being a means of contributing to the achievement of the just cause that involves intentionally attacking only those who are liable to attack. But precisely

1. This use of the term is unorthodox because it omits wars that are unjust for other reasons, e.g., because they are disproportionate. To appreciate why I limit the reference of the term for present purposes, see Saba Bazargan, “The Permissibility of Aiding and Abetting Unjust Wars,” *Journal of Moral Philosophy* (forthcoming).

because those against whom a just war is fought must be liable to attack on the basis of specific wrongs for which they are responsible, it is easy to see how a war that might promote some important aim, such as a reduction in the threat of terrorism, could nonetheless be unjust. For intentionally harming innocent people can sometimes be an effective means of achieving good aims, or what the agents regard as good aims. This is, indeed, the foundational assumption of terrorism. For this reason, the moral risks involved in fighting in a war that might be instrumental in the achievement of a good aim are greater than those involved in fighting in a particular battle in a just war.

The emphasis of Ryan's argument, however, is not so much that combatants cannot distinguish between just and unjust battles and thus must fight both; it is more that if they have the option of refusing to fight in those that are unjust, or that they believe or claim to believe are unjust, the military of which they are members can never function efficiently to win a just war. By the same logic, if soldiers in a democracy have the option of refusing to fight in a war that is unjust, or that they believe or claim to believe is unjust—that is, if their obedience cannot be guaranteed—then the military of which they are members cannot be confidently relied upon as an effective defender of the state, its people, and its democratic institutions.

I doubt that this objection requires the analogy between battles and wars, or the associated suggestion that even an unjust war may be seen as a part of a larger activity that is just overall. Ryan's point is simply that whenever combatants in a democratic state fight in obedience to legitimate orders, even in an unjust war, they thereby contribute to the enduring reliability of the institutions that protect their citizens and their democratic institutions.

Ryan's concern that the position I have defended would, if accepted, impair the ability of states to fight just wars is shared by Yitzhak Benbaji. It is, indeed, the principal concern that animates both of their critiques. The differences are matters of emphasis. Ryan contends that an obedient military is necessary in a democratic society to defend not only the state and its citizens but also their democratic institutions, "whose preservation and promotion potentially benefit everyone" (32). Benbaji argues that, given that states must rely on self-help in enforcing a restrictive doctrine of *jus ad bellum*, it is necessary for them to have obedient armies to defeat unjust aggressors and deter other potential aggressors.

Ryan is concerned with the problem of motivating soldiers in a democratic society to obey orders to fight in a just war. His main worry is that if soldiers in democratic societies are legally permitted and socially encouraged to engage in conscientious refusal when, after serious reflection, they believe a war is unjust, they might refuse to

fight in a just war of national self-defense, either because they mistakenly believe that the war is unjust or because they can exploit liberal allowances for conscientious objection as a means of avoiding the personal risks involved in fighting. This problem is arguably even more acute than Ryan recognizes. If legal and social norms in democratic states permit soldiers to disobey an order to fight in an unjust war, this may give a general advantage to unjust aggressors over just democracies. This is because nondemocratic states are more likely than democratic states to be tempted to engage in unjust aggression, to succeed in manipulating allegedly factual information that reaches their soldiers, to suppress advocacy of conscientious refusal, and to punish conscientious refusal savagely if it occurs. For these reasons, general encouragement of conscientious refusal to fight in unjust wars might on balance subvert rather than promote the aim of preventing unjust wars and might also make unjust wars more likely to succeed. For even if the promulgation of the view that it is wrong to fight in an unjust war may inhibit the initiation of unjust wars by democratic states, it may also weaken the deterrence of unjust wars by non-democratic states by compromising the ability of democratic states to fight in self-defense.

But these fears are exaggerated. Consider first the threat of mistaken beliefs. In almost all wars, at least one side fights unjustly—that is, fights a war that is unjust. There have therefore been a great many unjust wars. Yet the evidence indicates that comparatively few who fought in those unjust wars believed that their war was unjust. And among their adversaries who were fighting in a just war, even fewer believed that their war was unjust. People generally, and soldiers in particular, are deeply reluctant to accept that their country could be an unjust aggressor, or that they themselves are unjust aggressors. So in virtually every war in which some people fought unjustly, most of them fought in the mistaken belief that their unjust war was just. By contrast, how many instances are there in which there could have been a just war but it was not fought because of opposition on moral grounds from citizens or soldiers of the state with the just cause? How many times has a just war—in particular, a just war of national self-defense by a democracy—been lost because of misguided morally motivated opposition from the soldiers commanded to fight it? Or how many just wars *would* have been lost had soldiers not been compelled to fight by conscription enforced by draconian penalties for disobedience? How many instances are there of a democratic state ceasing to be a democracy because, for moral reasons, its soldiers resisted legitimate orders to defend it?

I assume that one will be hard pressed to find examples that answer to the descriptions given in these questions. One might argue

that this is just a consequence of the general acceptance throughout history of the view that soldiers do no wrong merely by fighting in an unjust war, so that if it became generally accepted that it *is* wrong to fight in an unjust war, just wars would begin to go unfought or be lost as a result of misplaced moral scruples. No doubt traditional moral beliefs are part of the explanation. But there are also deep roots in human psychology. People are naturally disposed to be trusting and loyal in evaluating the acts of their own political community, so that even if soldiers believed as a general matter that it is wrong to fight in an unjust war, the moral case against a war their government claimed was just would have to be unusually compelling to overcome their reluctance to accept that their leaders might be no better, or even worse, than their adversaries. The disposition to defer to the authority of one's own leaders is, moreover, reinforced by self-interest. Unjust wars of aggression tend to serve the interests of the aggressors, or at least the interests of the rulers and their cronies. In most cases, therefore, there are obvious incentives for a government that desires to fight an unjust war to use the resources at its disposal to persuade its citizens that the war would be just. But it is harder to identify incentives that a government, or anyone else within a political community other than those who are agents of a foreign power, might have to persuade the citizens that what would in fact be a just war of national self-defense would instead be an unjust war. It seems obvious that when what is in fact unjust aggression threatens not only one's own life but also the lives of one's loved ones and compatriots, as well as one's way of life and democratic institutions, it is improbable that one will conclude that military resistance would be unjust. Pacifists, of course, will draw that conclusion, but Ryan's principal concern is with soldiers, who tend not to be pacifists. (I do not mean to suggest that one cannot be mistaken about the permissibility of a defensive war. Wars of defense against justified humanitarian intervention are, in the absence of special circumstances, unjust.)<sup>2</sup>

While the probability is thus very low that soldiers will judge a just war of national self-defense to be unjust, the probability is higher that they might make that mistake in the case of a just war of collective defense or humanitarian intervention. There are various ways in which a war of humanitarian intervention might be unjust. It might, for example, be intended to stop some atrocity but only as part of a strategy of domination or conquest. Or it might be entirely altruistically motivated yet violate the rights of self-determination of the intended beneficiaries. Soldiers might, therefore, mistakenly believe

2. On this point, see Alex Leveringhaus, "The Moral Status of Combatants during Military Humanitarian Intervention," *Utilitas* (forthcoming).

that one of these objections applied to a particular instance of humanitarian intervention when in fact it did not. A democratic state's failure to fight a just war of humanitarian intervention would not, however, threaten its democratic institutions, which is Ryan's particular concern, though the failure to intervene might involve a failure to protect democracy elsewhere.

Just as it is improbable that soldiers would sincerely believe it unjust to fight in self-defense against genuinely unjust aggressors, so it is unlikely that they would pretend to believe this as a means of avoiding the personal risks involved in fighting. When soldiers who have been schooled in the ethos of a professional military are confronted with threats to their lives, their loved ones, and their way of life, not many look for pretexts to evade their professional duties. Yet Ryan cites various instances in which governments fighting what most people regard as just wars—the colonial government in the American Revolutionary War, the Union in the American Civil War, the French in response to the German invasion in the First World War—were compelled to propose or adopt a policy of conscription. But these examples are all problematic. The Revolutionary War was probably unjust on both sides. The grievances of the colonists against Britain did not rise to the level of a just cause, and a substantial proportion of the population opposed the war on reasonable grounds. But neither did the colonists' declaration of independence constitute a just cause for war by Britain. While the Union had a just cause for war—the abolition of slavery—it was not what principally motivated the government to resort to war. But to the extent that men in Union states perceived that the aim was to free slaves in the South, the war was from their perspective not so much a war of defense but a war to benefit strangers. Finally, while Ryan says that France “was compelled to institute conscription” (35), the French had in fact been practicing universal conscription since 1872, following the Franco-Prussian War. What happened in 1914 was only that the government mobilized its reserves and territorial militia, all of whom had been conscripted earlier. (France had a long history of conscription, beginning after the Revolution and continuing under Napoleon. It abandoned conscription only in 2001.)

I acknowledge that the temptation to evade the fighting is stronger if the war is not self-defensive, so that fighting would place one's life at risk for the sake of people to whom one is not specially related. It is therefore reasonable to suppose that in a democratic culture in which it was widely accepted that it is wrong to fight in the absence of a just cause and in which, as a consequence, there were generous provisions for conscientious objection, it would, as I conceded in *Killing in War*, be more difficult to fight just wars other than wars of

national self-defense, such as humanitarian wars.<sup>3</sup> But this would not, as I noted earlier, threaten domestic democracy. Moreover, most people believe that most humanitarian wars are supererogatory, or morally optional, at least when the expected costs for the intervening state would be high and soldiers would have a significant incentive to evade the fighting. If this is right, the likelihood is that any just humanitarian war that a democratic state might not fight that it would have fought had it retained the demand for unconditional obedience would be a war that it was morally permissible not to fight. The effects, then, of widespread recognition that it is wrong to fight in an unjust war are likely to be that, of those wars that *would* have been fought in the absence of this recognition, some that would have been impermissible would not be fought, a smaller number that it would have been permissible either to fight or not to fight would not be fought, and an even smaller number of obligatory wars would also not be fought. This would almost certainly be a significant improvement over the status quo. It would, of course, be even better if all those humanitarian interventions that would be justified (and therefore proportionate in their effects) were fought, even if they were supererogatory. But the problem of motivating people to fight just humanitarian wars is a distinct problem, even if it might be exacerbated by the introduction of liberal provisions for conscientious objection. It needs to be separately addressed. The best solution, as I suggested in *Killing in War*, probably lies in the establishment of an international force under international control whose sole function would be to conduct humanitarian interventions.<sup>4</sup> There are of course formidable obstacles to this, but we should address the problem directly rather than allowing it to inhibit our efforts to persuade people not to fight in unjust wars.

Ryan has other concerns about conscientious objection. For example, could a state realistically permit combatants to refuse on moral grounds to continue to fight even during combat operations? Probably not, but that is not a significant problem. The stressful and distracting conditions of combat are inimical to the careful deliberation necessary for reliable judgments about complicated moral issues. Even if a combatant experiences what he takes to be a moral epiphany in the midst of battle, he would be unwise to trust it. Only if it survives careful reflection later, in conditions more conducive to rational deliberation, should he conclude that he ought to stop fighting. Then, of course, the question arises whether a state could afford to allow combatants to retire from a combat zone for moral, or ostensibly

3. Jeff McMahan, *Killing in War* (Oxford: Clarendon, 2009), 100.

4. *Ibid.*, 100–101.

moral, reasons. Again, perhaps not, though retaining a sincere dissenter in a combat zone could be disruptive as well. In these conditions, a conscientious soldier may be morally required to suffer penalties for conscientious refusal that Ryan and I agree would be unfair. It may be that the best we can do is to try to prevent these situations from arising by trying to prevent unjust wars from occurring. And one way to do that, I have argued, is to encourage soldiers to refuse to fight in wars they can reasonably believe to be unjust.

Had the view for which I have argued been widely accepted during the Vietnam War, it might have significantly diminished the problem Ryan describes: that soldiers came to understand that the war was unjust only after they had begun to participate in it. If soldiers and potential conscripts had accepted that it was their responsibility to determine that the war was not unjust before allowing themselves to fight in it, more would have refused to go in the first place rather than having to discover the truth only after they had become complicit.

Ryan identifies an inconsistency between two claims I have made. One is that democratic procedures of the sort found in the United States do little to ensure that moral considerations are taken into account and given due weight in deliberations about the resort to war. The other is that “the military must not have the discretion to go to war on its own initiative” because it cannot be subject to the constraints that apply to democratic governments (38–39). In particular, “military decision-makers are neither chosen by the people nor representative of them” (39). I think these claims are not inconsistent, though my choice of wording is responsible for their appearing to be.

The procedural constraints of democratic decision making are valuable, not so much because they tend to produce morally defensible decisions but because they ensure, through the threat of expulsion from office, that those who make decisions about the resort to war are sensitive to the views of their citizens who may be put at risk by the war and must also provide the resources necessary to fight it. Democratic constraints may thus function to restrain a government from initiating an unjust war that would benefit an elite minority associated with the government but burden the majority. But it might also restrict the ability of the government to fight a just humanitarian war. In rare instances of the latter sort, I accept that it can be permissible for soldiers to fight in defiance of legitimate orders not to. That is, I accept that what Ryan calls “conscientious initiation” can be permissible, though only rarely. If, for example, there had been a contingent of U.S. forces among the peacekeepers in Rwanda in 1994 that had had the ability to use force to prevent the massacre of a significant number of innocent people, it would have been morally



permissible for them to do so despite the determined efforts of the Clinton administration to avoid U.S. involvement no matter how many innocent people might be butchered.

As Ryan notes, the areas of agreement between us are extensive. Despite his cogent exposition of the argument to democratic duty, he agrees with me that people ought not to fight in unjust wars. The difference between us is that he thinks that they ought not to fight in any wars, so that they ought not to become soldiers at all. He therefore takes the argument to democratic duty less seriously than I do. I accept the premise of that argument that it is necessary to “protect the protection” because we often have a duty to protect innocent people from harm inflicted by culpable threateners, whether foreign or domestic (a duty that seems to have been well fulfilled recently in Libya by several intervening states). Pacifists argue that we must get out of the business of protection, at least by military means. Traditional just war theorists and political realists claim, by contrast, that we must accept that an ordinary soldier does no wrong by fighting in an unjust war and must always obey a legitimate order to go to war. I defend the middle ground between the abandonment of forcible defense and the demand for blind obedience. I argue that soldiers ought to obey an order to fight in a just war but disobey an order to fight in an unjust war, and that this imposes a requirement on the rest of us to share the burdens of conscientious action with them, in part by enhancing their capacity to distinguish reliably between just and unjust wars. As Ryan’s forceful objections demonstrate, there are intractable problems in reconciling this position with the requirements of military efficiency. But given the enormity of all that is at stake, I think that we have no defensible alternative but to do our best to solve them.

## II. BENBAJI: CONTRACTARIANISM, LAW, AND MORALITY

Yitzhak Benbaji develops a complicated argument for the traditional view that it is morally permissible for unjust combatants to kill just combatants during a state of war. One of the premises, which he calls Mutual Benefit, rests on seven assumptions. The two on which I will focus are, first, that the rights that enable states to protect the rights of their citizens “can best be protected by . . . a prohibitive *jus ad bellum*, which condemns wars of aggression” (50), and, second, that this prohibitive regime can best be enforced only if states “are allowed to maintain obedient armies” (51). The overlap with Ryan’s argument is obvious: both insist on the necessity of obedience when soldiers are legitimately commanded to go to war—Ryan because toleration of disobedience imperils the ability of democracies to defend their in-

stitutions, and Benbaji because it threatens the ability of decent states to enforce the prohibition of aggression. Some argument of this sort almost certainly offers the best prospect of vindicating the central elements of the traditional theory of the just war. In his contribution to this symposium and in various other articles, Benbaji has pursued this core idea through the development of a remarkably intricate contractarian account of the morality of war.

He follows a hallowed tradition in just war theory and international law in supposing that the *ad bellum* norm that will best protect the rights of individuals is one that prohibits aggression. I think this is a mistake, at least if aggression is understood in the usual way as an attack by one state against another that has not itself attacked any other state. The problem is that there are wars that are aggressive in this ordinary sense that can nevertheless be morally justified, such as humanitarian interventions to protect the rights of people in other states, and wars to seize hoarded resources that are necessary for the survival of a people. There are also wars that are defensive and therefore not aggressive that are unjust, such as wars of defense against a justified humanitarian intervention or against a justified seizure of resources necessary for subsistence. At least as a matter of morality, a doctrine of *jus ad bellum* that prohibits certain just wars and permits certain unjust wars cannot be correct. But the claim that the correct doctrine prohibits wars that are unjust and unjustified but permits wars that are just or justified is merely formal and therefore trivial. For the purpose of discussing Benbaji's contractarian account of war and the ways in which it differs from the account I have defended, this does not matter. This discussion need not take a position on the substantive question of which wars are just, or justified.

Suppose it is right that our aim should not be "to prevent aggression by creating the optimal conditions for enforcing the prohibition on aggression" (61), but should instead be to prevent unusually serious wrongs, such as unjust wars, and to defeat those who fight such wars. To achieve these goals, which would be better: for states to have uniformly obedient armies or for them to have armies that obey orders to fight just wars but disobey orders to fight unjust wars? Obviously the latter, provided that wars come properly labeled as just or unjust, or that armies have the ability to distinguish infallibly between just and unjust wars. But they do not; therefore, as a practical matter, much depends on the reliability of soldiers' judgments, on the probability that they will correctly judge wars to be just or unjust. One possibility is that all soldiers tend to distrust their leaders and therefore assume that there is a presumption that any war in which they are ordered to fight is unjust. If that were true, governments disposed to fight unjust wars might be more successful in motivating

their armies to fight, since they would be likely to have fewer scruples about deceiving and coercing their own soldiers. In these conditions, a world of uniformly obedient armies would be better for the prevention of unjust wars than a world in which soldiers gave some weight to the deliverances of their own consciences and there were institutional provisions for their doing so.

But, as I argued in *Killing in War* and again in response to Ryan, soldiers are not like that.<sup>5</sup> They are disposed to trust the authority of their government and to accept that any war in which they are commanded to fight is presumptively just. They are therefore more likely to judge that an objectively just war is just than that it is unjust. They are, moreover, less likely in general to judge that a just war is unjust than to judge that an unjust war is unjust. In these conditions, which characterize the world as it is, the probability that the average soldier will judge a just war by his own state to be unjust is very low, and soldiers' judgments about whether their own state's wars are just or unjust are more likely to be correct than incorrect. In such conditions, the aim of preventing unjust wars will be better achieved if soldiers are reluctant to fight in wars they believe to be unjust than if armies are uniformly obedient. (As I write, an obedient army in the Syrian city of Hama is doing what the members of obedient armies often do: killing innocent people in an effort to reduce the survivors to a state of obedience as abject as their own.)

Benbaji responds to this by arguing that a government determined to fight an unjust war would respond to a convention that permits certain forms of conscientious objection by escalating its efforts to deceive and coerce its soldiers, thereby largely neutralizing any effect that the convention might otherwise have in preventing unjust wars. I have noted in reply that anything that increases the predictable cost of fighting an unjust war contributes to the deterrence of unjust wars. Benbaji replies in turn that such a convention would also require a government that wants to fight a *just* war to devote more resources to persuading its soldiers to fight. To illustrate this claim, he cites an example of the kind of war for which the justification is most likely to be obscure and least likely to be publicly demonstrable: preventive war. But one can concede his point without accepting that it establishes his case. For it takes more to deceive people into believing that an unjust war is just than it does to persuade people that a just war is just. This is a consequence both of the disposition to believe that one's own war is just and of the fact that deception requires going against the evidence while persuading people of the truth is supported by it. A convention that permits consci-

5. *Ibid.*, 108, 119–21.

entious objection would therefore increase the cost of fighting unjust wars by more than it would increase the cost of fighting just wars. This asymmetry supports the claim that such a convention would be better for preventing unjust wars than a convention requiring unconditional obedience.

Benbaji makes the further point that a convention that permits certain forms of conscientious objection would also weaken the ability of states faced with threats of unjust aggression to deter that aggression by credibly threatening to respond in a way that would be morally impermissible, for example, because it would be disproportionate. I accept this point and concede that a diminished capacity to bluff would be a loss. But there would also be a corresponding moral gain: namely, that the convention would also diminish the risk that a threat of impermissible retaliation would actually be carried out if it failed in its deterrent function.

In summary, Benbaji is right that, relative to a convention that demands unconditional obedience, one that permits conscientious objection would make it harder to fight just wars, especially just wars that are not self-defensive in nature, and would weaken the ability of states to deter unjust aggression by threatening indiscriminate or disproportionate retaliation. These are genuine costs. What I have tried to show is that these costs would be substantially outweighed, particularly by the convention's effect in making it more difficult for states to initiate unjust wars.

Another premise of Benbaji's argument, Consent, states that by accepting the role of a soldier, one consents to waive one's right not to be killed by enemy combatants. Together with the other two premises, this is supposed to yield the proposition he calls Waiver, which states that this consent by soldiers is "morally effective"—that is, that it makes it morally permissible for their enemies to kill them. I argued in *Killing in War* that if this were true, one might reasonably refuse to grant unjust aggressors that permission by fighting not as a soldier but simply as an individual.<sup>6</sup> Benbaji agrees that this is possible. "Contractarianism," he writes, "does not deny that individuals have a natural (or pre-conventional) right to defend themselves: individuals are at liberty to fight as partisans" (59). Notice, however, what this entails: that the potential victims of unjust aggressors have the power to determine whether the aggressors act permissibly or impermissibly in killing them. If the potential victims choose to resist by becoming soldiers in a legally recognized military organization, it is permissible for the aggressors to kill them. But if instead they resist by organizing a *levée en masse*, it is impermissible for the aggressors to kill them. If,

6. *Ibid.*, 55.

for example, unjust aggressors invade Nicaragua and are opposed by soldiers of the Nicaraguan Army, they do no wrong in killing them; but if they instead invade neighboring Costa Rica, which has no army, and meet with coordinated defense by people who take up arms as individuals, their acts of killing are all impermissible. Benbaji's view must concede that in the distant past, before there were states and professional armies, all unjust warriors were murderers or wrongdoers. The moral progress we have made since then, according to his view, is that we have adopted conventions that have made what would once have been murders permissible.

Suppose, however, that Benbaji is right that soldiers effectively waive rights they would retain if they fought as individuals, so that enemy combatants do not wrong them when they kill them. Next suppose that unjust aggressors invade a person's state and that, because he has neither weapons nor training in fighting, the only way he can contribute to the just defense is by joining the army. If he is morally required to participate in the defense, the aggressors will, through their wrongful action, have compelled him to waive his right that they not kill him. Even if they do not violate his right when they kill him, they will nonetheless have *wronged* him by having wrongfully created the conditions in which it became permissible for them to kill him. (Or, if he joined the military prior to the invasion, they have wrongfully exploited his earlier waiver.) This cannot be said of just combatants who kill unjust combatants, for even if the unjust combatants had not earlier waived their rights, they would have forfeited them prior to being killed. Thus, even if we grant Benbaji's claim that all combatants waive their rights, there is still a profound moral asymmetry between just and unjust combatants.

It is worth mentioning, if only parenthetically, a further, more important asymmetry. As pacifists such as Ryan rightly emphasize, war almost inevitably involves the killing of innocent bystanders as a side effect. Whereas just combatants often have a necessity justification for harms their military action foreseeably inflicts on innocent bystanders, this form of justification seldom, if ever, applies to the action of unjust combatants. Furthermore, to the extent that an unjust war is successful, unjust combatants contribute to the wrongful harms inflicted on the victims of the achievement of their state's unjust aims. But there is no corresponding objection to the contribution that just combatants may make to the achievement of a just cause. Benbaji acknowledges the first of these differences. Although he says that spatial constraints prevent him from considering it, he also hints that it might be addressed by noting the way in which civilians are "part of the contractarian scheme" (49). But it seems highly unlikely that it can be shown that civilians in a state that is the victim of wrongful

aggression consent to be killed as a side effect of an aggressor's military action, or that they consent to surrender whatever the aggressors succeed in depriving them of. If this is right, the contractarian approach cannot vindicate the claim of full moral symmetry between just and unjust combatants even if it succeeds in showing that unjust combatants act permissibly when they intentionally attack and kill just combatants.

I now proceed to what I think is the most damaging objection to Benbaji's contractarian account of war rights. He writes early in his essay that "McMahan denies that symmetry is mutually beneficial: a regime under which soldiers have no legal right to participate in a war of aggression is better for decent states . . . than a regime which allows obedience" (47). This is actually not what I argued in *Killing in War*, though I have speculated elsewhere that at present it would be best if the law made participation in an unjust war illegal though not criminal—that is, legally condemnable but not punishable.<sup>7</sup> My primary concern has instead been with *moral* rights: whether just combatants retain their moral right not to be killed and, if so, whether unjust combatants could have a moral right to kill them. Yet Benbaji's reference is to *legal* rights, which remain the focus of his argument throughout. His basic claims are these: that the legal rights of just and unjust combatants are the same, that neither type of combatant has a legal right not to be killed by the other, that this allocation of legal rights is acceptable because it is fair and mutually beneficial, and that this understanding of the legal rights of a soldier is inherent in the role of the soldier, so that when a person becomes a soldier, he or she accepts this role and thus consents to having no legal right not to be killed by enemy combatants. When, for example, Benbaji defends his claim that, "by signing up, soldiers waive their right against being unjustly attacked," what he says is that "the authority of states to require obedience has never been seriously challenged in the international community. Therefore, it must be widely acknowledged that the law denies soldiers a legal right not to be unjustly attacked, which in turn supports the claim that soldiers themselves share this understanding of their role" (62). "This understanding," however, is that what people consent to give up by becoming soldiers is "their legal right not to be unjustly attacked," *not* their moral right not to be attacked or killed (62).

Yet Benbaji claims that "contractarianism shows how legal rights, conferred on soldiers by a fair and mutually beneficial institutional scheme, become moral rights of those who are governed by this scheme" (49), and he also believes, more importantly, that when a

7. *Ibid.*, 105–10, 189–92.

fair and mutually beneficial legal scheme *denies* soldiers certain legal rights, the absence of those legal rights becomes the absence of corresponding moral rights. But he never explains the process by which this transmutation of the legal into the moral occurs. It is clearly true that all soldiers lack a legal right not to be killed by enemy soldiers during a state of war and that virtually all soldiers are aware of this and accept it—or, if one prefers to express it this way, that on becoming soldiers they consent to be without a legal right not to be killed by enemy soldiers. Even if it is also true that this legal arrangement is fair and mutually beneficial, there is nothing here to support the claim that soldiers waive their *moral* right not to be killed, so that it then becomes morally permissible to kill them. If people rightly perceive it to be in their interest to accept a legal system that denies legal protection to certain of their moral rights, it may well be morally permissible for them to institute that legal system.<sup>8</sup> In *Killing in War*, I argue that this is precisely what we have done in the case of the law of war and that this explains why the law diverges so radically from the morality of war.<sup>9</sup> But the fact that we find it mutually beneficial and fair to deny ourselves certain legal protections has no bearing on what moral rights we have. Consenting, for whatever reason, to do without a certain legal protection in no way entails the waiving of the corresponding moral right.

This unwarranted inference of moral rights from legal rights is also found in Benbaji's discussion of marriage. He observes, rightly, that the legal rights of marriage are determined by the nature of the institution, not by what married people believe about those rights. Even if a married couple have always believed that they have no option of divorce, they nevertheless have that legal option. Benbaji says that "their legal right of exit is created by their consent to become married" (60). But consent may have nothing to do with it. Suppose the husband was systematically deceived, so that during the marriage ceremony he believed that he was undergoing a mysterious ritual to become a Freemason. He did not then consent to the terms of marriage as a legal institution. But if his marriage was legally valid, he has the legal rights of a married person. (Benbaji also claims, by reference to the analogy with marriage, that a person can lose his moral right not to be killed by consenting to be a soldier even if he does not understand what he is consenting to. But suppose a person goes through the process of joining the military in the belief that he is

8. I accept that it is; David Rodin believes that it is not. See his "Morality and Law in War," in *The Changing Character of War*, ed. Hew Strachan and Sibylle Sheipers (Oxford: Oxford University Press, 2011).

9. McMahan, *Killing in War*, 105–10.

joining the Freemasons. He may then lack a legal right not to be killed by enemy soldiers, but he has not waived his moral right not to be killed.)

That the couple in Benbaji's example have the legal right of divorce does not entail that they have a moral right of divorce. Many people believe that marriage is a moral state as well as a legal one and that there are moral dimensions to marriage that are independent of the legal features of that state. By virtue of their shared understanding of the married state, the people in Benbaji's example may well have denied themselves a moral right of divorce when they entered the married state. That they uncontroversially have the legal right is insufficient for their also having the moral right.

Although the complexity of Benbaji's account makes it difficult to determine the precise architecture of his argument, the process by which it purports to convert the legal and conventional into the moral not only seems alchemical in nature but is also deeply conservative in tendency. For any accepted practice, one can first argue that its terms are fair and mutually beneficial for those who participate in it, then argue that those who have institutional roles in the practice consent to those terms when they adopt their role, and—presto!—the terms of the practice become moral rights and duties. The workings of the conversion process are perhaps best visible in the section of Benbaji's essay on transferred responsibility. There he contends that the social and legal conceptions of the role of a soldier to which a person consents by joining the military include the idea that soldiers waive their legal right not to be killed vis-à-vis enemy soldiers but not vis-à-vis enemy governments. So unjust governments violate the rights of just combatants though unjust combatants do not, *provided that* in their contract with one another, "states undertake the duty to make sure that the wars they fight are just" (64), which of course they do. The label "nuanced contractarianism" is fitting here, for this account of what soldiers consent to is nuanced indeed. What Benbaji has packed into the role of the soldier is nothing less than the basic elements of the traditional account of the just war defended by Walzer and instantiated in the international law of war: namely, that *jus ad bellum* applies only to governments, not soldiers; that only *jus in bello* applies to soldiers; and hence only governments are responsible for the killing of just combatants, since they are not protected by the principles of *jus in bello* but only by the principles of *jus ad bellum*. The claims about moral rights in Benbaji's conclusion are all there in the form of true claims about legal rights in the premises. But, again, even if the distribution of legal rights is fair and mutually beneficial, and people adopt the roles from which certain legal rights are excluded,



it does not follow from their having waived those legal rights that they have also effectively waived the corresponding moral rights.

### III. RODIN: LIABILITY, NARROW PROPORTIONALITY, AND NUMBERS

As will be apparent to readers of David Rodin's article, I am in broad sympathy with his ambitious arguments to show that both liability and lesser evil justifications for the infliction of harm are thoroughly suffused with proportionality judgments. There are, nonetheless, many small disagreements between us. But rather than discuss those, I will focus primarily on one point of disagreement that may be of greater significance. It concerns the ways in which the number of people harmed by defensive action can affect whether there is a liability justification for that action.

My thinking about the relevance of numbers to narrow proportionality—that is, proportionality in the infliction of harm on people who are potentially liable to be harmed—has evolved, though not nearly enough, since I wrote *Killing in War*, so that I now think the two passages on this issue that Rodin quotes from the book are mistaken. First, it is false that harms to which people are liable do not count in any proportionality calculation. They count in the determination of narrow proportionality. My earlier thought was just that because their infliction is justified by the victim's liability, they do not count against the act of defense. But, as I will indicate later, I now suspect that the truth may be more subtle than this. Second, it is misleading to say, as I did, that “harms to which people are liable” are “discounted in proportionality calculations.” That is the view I took in much earlier work, and it is the view that Rodin takes of the way certain harms can count against the permissibility of action, but I am now skeptical of the idea that harms to which people are liable are discounted in any form of proportionality assessment.<sup>10</sup>

To introduce the relevant problem, it will help to compare two simple hypothetical examples.

1. Each of 1,000 culpable threateners will kill me unless I kill him.
2. Each of 1,000 innocent threateners will kill me unless I kill him.

In both cases, I can kill all 1,000 threateners through a single act of self-defense. A “culpable threatener” is someone who acts impermis-

10. See Jeff McMahan and Robert McKim, “The Just War and the Gulf War,” *Canadian Journal of Philosophy* 23 (1993): 501–41.

sibly in what Parfit calls the “fact-relative” sense in threatening an innocent person and to whose action no excusing or even mitigating conditions apply. An “innocent threatener” is someone who is morally responsible for a threat of wrongful harm to an innocent person’s life but is not culpable. There are various ways in which one might be responsible but not culpable. One may choose to act when it is reasonably foreseeable that one will impose a risk on others but the risk is sufficiently slight that one’s action is permissible in what Parfit calls the “evidence-relative” sense.<sup>11</sup> Or one may act under duress that is fully exculpating. (There are types of threatener intermediate between the innocent and the fully culpable, but they are irrelevant for present purposes.)

Many people, perhaps most, believe that it is permissible to kill all 1,000 culpable threateners. If there were only one, he would be liable to be killed by me in self-defense. And that is true of all 1,000: there is a liability justification for killing each one. The numbers seem not to matter, at least to many people. As Rodin says, the harms “are not aggregated, but considered separately” (99). But intuitively, the case of the innocent threateners seems different. According to the account of liability I have defended, if there were only one innocent threatener, I would have a liability justification for killing him. But in this case, the numbers seem to matter, or to matter more. As the number who would have to be killed to preserve my life increases, it seems intuitively that a point must be reached, probably before there are 1,000, at which it becomes impermissible to kill them all in my defense.

Rodin offers an explanation of why it may be impermissible to kill all 1,000 innocent threateners in self-defense, though the terms in which he describes it obscure rather than clarify it. Here is what he says. Because a liability justification considers the relation between the potential victim and each threatener separately, and because each innocent threatener is liable to be killed, there are liability justifications for killing all 1,000. But a “lesser evil justification . . . aggregates the defensive harms” (99). Although “it discounts the evil attributed to harm inflicted on the liable . . . unless the harm is discounted to zero, it is still possible that defensive harm inflicted on multiple liable persons will,” when aggregated, “not be the lesser evil” (99). Rather, it may be that the aggregated harms that would be inflicted on all 1,000 innocent threateners make “self-sacrifice the obligatory lesser evil” (100).

It is difficult to make sense of this. Rodin is ostensibly contrasting a liability justification with a lesser evil justification. But if there is a

11. Derek Parfit, *On What Matters* (Oxford: Oxford University Press, 2011), 1:150–64.

liability justification for killing the 1,000, the absence of a lesser evil justification for killing them is irrelevant. Nor can he mean that there is a lesser evil justification for self-sacrifice. For self-sacrifice would seem to be supererogatory; if it is, it is permissible and does not require a justification. So what does he mean by the “obligatory lesser evil”?

I think he means that there is what might be called a *greater evil objection* to killing all 1,000. Although the harm to each is discounted for his moral responsibility for a threat of wrongful killing, that harm remains an evil. These harms, albeit discounted, add up the more threateners there are, until a point is reached at which they together outweigh considerations of liability. The deontological liability-based justification for harming is overridden by the aggregate bad consequences that acting on it would have, making it impermissible to act on the justification. This is the reverse of the more familiar phenomenon of a deontological *constraint* being overridden by the bad consequences that *respecting* it would have, making it permissible to violate the constraint.

Rodin says that this way of explaining the relevance of numbers is an instance of “consequentialist reasoning,” but strictly speaking it is not. Consequentialism is maximizing, yet Rodin does not think that a liability justification for an act of harming is overridden whenever the consequences of acting on it, appropriately discounted, would be worse than the consequences of some alternative act. He also rightly accepts that it matters to a lesser evil justification whether one inflicts the harm or merely allows it to occur, whether one causes or allows it to occur intentionally as a means or unintentionally as a side effect, and so on. But these are distinctions characteristic of deontological rather than consequentialist reasoning.

Rodin also suggests that the reason that it might be wrong to kill all 1,000 innocent threateners is a matter of “proportionality in lesser evil justification (wide proportionality)” (99). But again, this is not so. This is in part because what is at issue is not a lesser evil *justification* but a greater evil *constraint*. And there is also no issue of wide proportionality. Wide proportionality is by definition concerned only with harms caused to those who are not liable; yet the innocent threateners are by hypothesis individually liable to be killed. Because they are liable, killing them cannot be disproportionate in the narrow sense. Rodin’s explanation of the impermissibility of killing them is, therefore, not a matter of either wide or narrow disproportionality as I understand those categories.

Although Rodin thus misdescribes his view, his core idea seems sound: liability justifications can be overridden by considerations of consequences. Consider the example he cites as an embarrassment

for my view, but on which I think my view delivers a plausible verdict—namely, the case of a conscientious driver whose car unexpectedly goes out of control and will kill an innocent pedestrian unless the pedestrian destroys the car, thereby killing the driver. The driver made a choice to pursue her interests in a way that exposed the pedestrian to a tiny risk of death. Through mere bad luck, her choice has now made it the case that either she or the pedestrian must die. The pedestrian bears no responsibility for this situation. If other things are equal, the driver, not the pedestrian, should bear the costs of her own choice to engage in a risk-imposing activity. But other things may not be equal. The driver may be twenty years old with a long and happy life in prospect, while the pedestrian may be ninety-five with less than a month to live in any case. In these circumstances, the pedestrian's liability justification for killing the driver in self-defense is arguably overridden.

But is this the right explanation of why it may be impermissible to kill all 1,000 innocent threateners? The problem is that the reason it gives for thinking that the liability justification is overridden in that case applies also in the case of the 1,000 culpable threateners. The only difference is that Rodin might argue that the discount rate for harms to culpable threateners is steeper, in which case the number of culpable threateners that it would be permissible to kill before the liability justification was overridden would be higher. But still there must be a point at which the number of thoroughly villainous threateners who would kill me if I did not kill them would be high enough that I would be morally required to allow them to kill me. In practice this is not implausible because even villains have friends, relatives, employers, and so on who are innocent and would be harmed as a side effect of their being killed, so that the side effects of killing a large number of them in self-defense would make the killings disproportionate in the *wide* sense. But would the harms that the culpable threateners themselves would suffer really add up to make it impermissible for an innocent person to defend his life against them? Intuitions differ, but it is relatively common in the philosophical and legal literature for writers to claim that there is no limit to the number of culpable threateners an innocent person may kill if that is necessary to prevent them from killing her.<sup>12</sup>

As Rodin notes, in *Killing in War*, I offered an explanation of the relevance of numbers to narrow proportionality that appealed to the

12. This view is also reflected in the traditional view of *jus in bello*. As Thomas Hurka notes, "*In bello* proportionality as standardly understood seems to allow a nation to kill virtually any number of enemy soldiers to save just one of its own soldiers" ("Proportionality in the Morality of War," *Philosophy and Public Affairs* 33 [2005]: 34–66, at 58).

fact that in some cases the more people there are who contribute causally to the production of a harmful outcome, the less important each one's contribution is. The degree of each one's liability is therefore diminished to the extent that defensive action against him will do less to avert the threatened harm.<sup>13</sup> But this explanation is of limited application. It does not, for example, explain why it might be disproportionate in the narrow sense to kill all 1,000 innocent threateners, since each one of them will kill me unless I kill him.

There may, however, be another way in which numbers can affect a liability justification. In most situations in which it is unavoidable that some person or persons must be harmed, or perhaps in all such situations, there is in principle some distribution of unavoidable harm among potential victims that would be ideally just. In the ideally just distribution, each person takes his or her fair share of the harm. A liability justification for harming allocates unavoidable harms in the way that best approximates the ideally just distribution. Sometimes the ideally just distribution is a practical possibility, but more often it is not. In these latter cases, the amount of harm to which a person is liable may depend on the circumstances. Suppose, for example, that there is a fixed harm that either you must suffer or I must suffer. You bear most of the responsibility—say, 90 percent—for our predicament, though I bear the remaining 10 percent. Assuming there are no other factors relevant to the just distribution of the harm, the ideally just distribution would be for you to suffer 90 percent of it while I suffered 10 percent. If it were possible to distribute the harm in that way, you would be liable to suffer 90 percent. But since the entire harm must go to one of us, the nearest approximation to the ideal is for it to go to you. In the circumstances, you are liable to suffer the entire harm. Yet even though there was a liability justification for inflicting the harm on you, there is a residual injustice. You have suffered more than your fair share. If it later becomes possible for me to compensate you for your having suffered the 10 percent that ideally I ought to have suffered, it is plausible to suppose that I owe you that compensation as a matter of corrective justice.

Return now to the case of the 1,000 innocent threateners. If there were only one such person and killing him were necessary to prevent him from killing me, he would be liable to be killed. But because he is not culpable, the outcome in which he gets all the harm and I get none may not be ideally just. Even though I bear no responsibility for our predicament, I ought, if possible, to accept a certain amount of harm to avoid killing him. If, for example, I could prevent him from killing me by merely wounding him, though this would require me

13. McMahan, *Killing in War*, 23–24.

to suffer a significant harm, such as the loss of a finger, it might be that he would be liable only to be wounded, not to be killed, and I would have to accept, though I would not be liable to, the loss of a finger. For that would be my fair share.

If there were only one innocent threatener, the harm he would suffer beyond his fair share is far less than that which I would suffer beyond my fair share if he were to kill me. While the latter remains constant even as the number of innocent threateners I kill in self-defense increases, the harms they together suffer beyond their fair share add up. I suggest that these are what weigh against—and, when the numbers get sufficiently high, outweigh—the harm I would suffer in being wrongly killed. While Rodin's explanation of why it might be impermissible to kill the 1,000 innocent threateners is simply that sufficiently bad consequences can override a liability justification, I suggest that the explanation is instead a matter of comparative injustice—that is, the harms the innocent threateners would suffer beyond their fair share together involve a greater injustice than the one I would suffer in being killed.

In a recent paper in which I gestured toward an explanation of this sort, I suggested that it offers a way to distinguish between the defensive killing of the 1,000 innocent threateners and the killing of 1,000 culpable threateners.<sup>14</sup> The difference is that the killing of a culpable threatener involves no residual injustice, as the ideally just distribution is that in which all the harm goes to him and none to the innocent potential victim. But that was a mistake. An innocent potential victim ought to accept *some* small harm if that would be the cost of defending herself by nonlethal rather than lethal means.<sup>15</sup> But if this is the case, there are at least small residual injustices even when culpable threateners are killed in self-defense. They are, like the discounted harms in Rodin's view, smaller than the corresponding costs involved in killing innocent threateners, but they add up nonetheless, so that it seems that there must be a limit to the number of culpable threateners that it is permissible for an innocent person to kill in self-defense.

But perhaps this does not follow. The degree of harm that an innocent potential victim might be required to suffer in order to defend her life against a fully culpable threatener by nonlethal rather than lethal means is presumably quite small. It is also plausible to

14. Jeff McMahan, "Who Is Morally Liable to Be Killed in War," *Analysis* 71 (2011): 544–59.

15. There are questions about the determination of people's fair shares of unavoidable harm that I cannot address here. Much depends, e.g., on whether the possible distribution of harm is assumed to be zero sum.

suppose that, for harms below a certain threshold of badness (the *threshold of additivity*), there is no number of such harms that, if suffered by different persons, could together be worse than, or outweigh, the death of an ordinary innocent person.<sup>16</sup> If, for example, one could either prevent each of a number of people from suffering a moment's mild pain or prevent the death of one innocent person, there may be *no* number of people whose mild pain it would be better to prevent than to prevent the single death. Finally, it is possible that any harm that an innocent potential victim might be required to bear in order to avoid having to kill a culpable threatener in self-defense would be below the threshold of additivity. If so, that would mean that in the case of the 1,000 culpable threateners, the harm that each threatener would suffer in being killed that would be beyond his fair share would be below the threshold of additivity. Those unfair burdens could then never outweigh the unjust harm that the innocent victim would suffer in being killed. That would be true no matter how many culpable threateners the victim had to kill in self-defense.

By contrast, it is reasonable to assume that an innocent person might be required to bear a harm beyond the threshold of additivity in order to defend her life against an innocent threatener by non-lethal rather than lethal means. If so, there could be a number of innocent threateners sufficiently large that the harms beyond their fair share that they would suffer in being killed would together outweigh the unjust harm that an innocent person would suffer in being wrongly killed. That would explain why it might be wrong to kill all 1,000 innocent threateners.

I concede that this explanation is speculative. Perhaps Rodin's simpler proposal is more plausible. But the explanation I have sketched better captures my sense that the objection to killing 1,000 innocent threateners is not just that it has consequences sufficiently bad to override a liability justification but that it wrongs the victims in a way that involves a greater injustice than that which is done to the one innocent victim when he is killed. According to this explanation, but not according to Rodin's, the impermissibility of killing the 1,000 innocent threateners is a matter of narrow disproportionality.

One other important issue that Rodin discusses is whether there are restrictions on the good effects that count in determining whether an act is proportionate in the wide sense—that is, in its effects on people who are not liable to be harmed. As a counterexample to the

16. For a much subtler discussion and defense of views of this kind, see Larry S. Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (New York: Oxford University Press, forthcoming), chap. 2, esp. secs. 2.2 and 2.3 and the references cited there.

suggestion that all good effects count, he cites a case in which, in order to fulfill a promise to his wife to be home by 6:00, a surgeon must drive recklessly. Suppose he does so, accidentally kills an innocent pedestrian, but is then able, as he earlier foresaw, to use one of the pedestrian's organs for transplantation into the body of a patient who would otherwise have died. If all good side effects of an act count, it seems that the surgeon's reckless driving was proportionate in the wide sense, for saving the patient counterbalances the killing of the pedestrian, leaving the intended effect—the fulfillment of the promise—as a net good effect. Rodin claims, and I agree, that this is a mistaken way to assess wide proportionality in this case. I am nevertheless skeptical of the principle he proposes to distinguish good effects that count from those that do not—namely, “only the beneficial consequences of the intended objective of action can offset the unintended harms for proportionality. The beneficial consequences of the unintended harm cannot offset the unintended harm” (102–3). This is similar to Thomas Hurka's earlier suggestion, with respect to proportionality in war, that good effects count if they are consequences of the achievement of the war's just cause but not if they are “side effects of the process of achieving” the just cause.<sup>17</sup> The difference is that Hurka's proposal is more restrictive, since consequences of unintended harms are only a subset of the consequences of a process of achieving a just cause.

The point on which Rodin and Hurka agree—that good effects count when they are further consequences of the achievement of the intended good aim (which in a just war is the just cause)—has considerable *prima facie* plausibility. This is in part because these good effects are often ones that have been absent because of the wrongful action of those who are liable to be harmed. Suppose, for example, that the achievement of the just cause of overthrowing a dictator has as a consequence a renewed flourishing of the arts in the liberated country. This good effect certainly seems to count if the reason the arts languished before the overthrow was that they were being suppressed by the dictator. If, however, the arts begin to flourish only because the war itself stimulated the imaginations of poets, novelists, and painters, then it may seem that this good effect does not count in the assessment of wide proportionality.

Despite its intuitive appeal, I am skeptical of the principle that Rodin proposes. A variation of Rodin's own example provides a counterexample. Suppose that if the surgeon speeds to get home by 6:00, his wife will interpret this as confirmation that he loves her and will then refuse to leave him, predictably prompting an unstable aspiring

17. Hurka, “Proportionality in the Morality of War,” 43.



lover to kill himself in despair, thereby making his organs available for life-saving transplants for two patients. If the surgeon then accidentally kills a pedestrian in his haste to get home, the two lives that will be saved by the transplants *do* count, on Rodin's principle, because they are consequences of "the intended objective of the action"—namely, the surgeon's arrival by 6:00. The reckless driving is then proportionate because the saved lives of the patients offset the deaths of the pedestrian and the disappointed lover, again leaving the intended fulfillment of the promise as a net good effect. Yet the surgeon's action in this variant seems no more defensible than it is in the original.<sup>18</sup>

I do not have a unifying criterion for distinguishing between good effects that count toward wide proportionality and those that do not.<sup>19</sup> There are, I suspect, numerous reasons why some good effects may not count, or may count but have lesser weight in some contexts than they have in others—for example, if they are caused in certain ways rather than others. Rodin is right, I think, about many of the details. It is, for example, plausible to suppose that when a bad side effect is a means to a further good effect, that good effect either does not count or has less weight than it would if it were a side effect of the act itself. (This is true even when, as in the variant of Rodin's case, both the bad side effect and the further good effect are consequences of the achievement of the intended good aim rather than of the means to the achievement of that aim.)

Another important factor—one that is part of the explanation of why the surgeon's action in Rodin's original case is not proportionate in the wide sense—is the moral asymmetry between doing harm and allowing harm to occur. That an act has as a side effect the saving of one person cannot fully offset its having the killing of a different person as another side effect. A related but distinct factor is the moral asymmetry between benefiting and harming. Assume for the sake of argument that there are objective measures of the magnitudes of at least some benefits and harms, so that experiencing a benefit of a certain magnitude would fully and objectively compensate a person for suffering a harm of that same magnitude. It does not follow that, in a determination of wide proportionality, the conferral of a benefit on one person as a side effect would fully offset the infliction of a harm of the same magnitude, or even a somewhat lesser magnitude, on a *different* person as a side effect.

A further kind of good effect that seems not to count in deter-

18. Assuming that proportionality is an objective matter, nothing hinges on whether the despairing lover's suicide is predictable.

19. I do, however, say a great deal more about these issues in a lengthy unpublished manuscript called "Proportionality in Self-Defense and War."

mining wide proportionality is a benefit to a wrongdoer who, in the circumstances, is liable to be harmed. Suppose, for example, that an act of war by just combatants would both kill some innocent bystanders and, for some reason I will not even attempt to describe, greatly enhance the material wealth of many of the unjust combatants against whom they are fighting. Intuitively, it seems clear that if the act of war would be disproportionate if its only side effects were the deaths of the innocent bystanders, it could not become proportionate by factoring in the beneficial side effects for the unjust combatants. This example also shows that the reasons why certain different good effects do not count may be quite heterogeneous.

There are various other disagreements that I wish I could discuss here. I think, for example, that death is in principle compensable, that probabilities of death are relevant to the determination of liability (for if the infliction of a nonlethal harm would be proportionate and being exposed to a certain risk of death would be less bad than a certainty of that nonlethal harm, then the imposition of that risk of death should be proportionate as well), that imminence and causal immediacy are irrelevant except as proxies for probability and reasonable foreseeability, that a lesser evil justification exempts an agent from liability to defensive action though not from liability to pay compensation, and so on. But spatial limits prevent me from taking up Rodin's forceful challenges on these points.

#### IV. GARDNER AND TANGUAY-RENAUD: DESERT AND DEFENSE

In various writings over the years, I have argued against the relevance of desert to the justification of either individual self- or other-defense, or defensive harming and killing in war. I have argued, for example, that a person can be liable to be harmed or killed in self-defense on the basis of his moral responsibility for a threat of wrongful harm even if he is not culpable, perhaps because he permissibly chose to risk becoming a threat or because he has a full excuse for posing a threat. Any account of defense that insists that a person must deserve to be harmed in order for it to be permissible to harm him in self-defense is therefore excessively restrictive. (I accept, however, that culpable responsibility for a threat is a stronger ground of liability to defensive action than responsibility without culpability; so culpability is relevant to the justification of self-defense even if desert is not.)

I have also argued that, because defensive harming is necessarily instrumental, the infliction of harm cannot be justified for defensive reasons unless doing so is to some extent instrumentally effective in averting a threat. The justification for defensive harming therefore cannot appeal to a consideration that provides a reason for harming

a person that is independent of its effect in averting a threat. Yet desert is such a consideration. If the justification for harming a person who poses a threat is that he deserves to be harmed, there seems to be a reason for harming him even if doing so would neither avert nor diminish the threat he poses. But if there is such a reason, it is not a reason for defensive action. This is what Gardner and Tanguay-Renaud (to whom, for the sake of brevity, I will often refer as “the authors”) call the “avoidability argument.”

They explore the possibility that desert might have a role, albeit a limited one, in the justification of defense. If successful, their efforts would yield a substantively restrictive account of permissible self- and other-defense that prohibits defensive harming of people who pose a threat but are not culpable for doing so. They would also forge a link between the justification of self-defense and the justification of punishment. The assertion of such a link is increasingly common, though the usual claim, as the authors note, is that an account of punishment can be extracted from the principles governing self-defense. This ambition to justify a practice of punishment with as little reliance on claims about desert as possible is one to which I am sympathetic. But they suggest the possibility of reversing this logic, grounding defense in part on considerations and principles that are widely thought to justify punishment.

They develop subtle and ingenious challenges to the avoidability argument and, in so doing, mount a formidable defense of the role of desert in the justification of self-defense. Their first claim is that there is a plausible and widely accepted account of the justification of punishment according to which punishment is not justified unless it satisfies both of two conditions—namely, that the person punished must deserve to be punished by virtue of having culpably engaged in wrongdoing, and that the punishment must be instrumental to the achievement of some good, such as defense, deterrence, or rehabilitation, that is distinct from the person’s getting what he deserves. But this is, in formal terms, exactly the form of justification that I reject in the case of defense.

It may seem to be essential to a justification with this structure that desert functions only as a necessary condition of permissibility rather than as a factor with positive, reason-giving normative force; for otherwise it seems that desert alone could justify punishment even when it would produce no further good—that is, when punitive harming would be “morally avoidable.” This is indeed their preferred understanding of the way in which desert functions in the justification. The person’s desert may simply negate the standing duty not to harm him without itself providing any reason to harm him. It may simply remove a moral barrier to action based on other, independent rea-

sons, such as reasons of defense. Hence, the authors deny that their principle NCJ1 gives “reasons . . . in favor of self-defensive actions. . . . That is because it does not give any reasons to do anything. . . . It is purely permissive” (126).

This is a coherent proposal, but it comes with certain costs, particularly since they claim that NCJ1 not only governs the morality of defense but “regulates intentional inflictions of harm in general” (119). For NCJ1 states a demanding condition of permissible harming—that is, that “it is morally permissible intentionally to inflict suffering or deprivation (. . . “harm”. . .) only on those who deserve such an infliction” (117). The authors point to one implication of the principle: “Suppose D defends herself excessively by the light of NCJ1 but does so in a reasonable misapprehension of what E is about to do to her. The reasonableness of her misapprehension would make her not guilty for the purposes of NCJ2, and that would rule out E’s having a permissible self-defensive response” (119). This may seem plausible, but there are other implications that are less congenial. First, suppose that a culpable aggressor reasonably believes that his victim’s defensive response is excessive and thus defends himself in a way that would be permissible if his belief were true. His misapprehension would make him not guilty and thus would rule out the victim’s having a permissible defensive response, since the desert condition would not be met.

Second, if a person’s deserving to be harmed does not itself provide a reason to harm him, but merely enables other reasons to harm him to become normatively effective, one must ask whether the same is true of a person’s deserving to be benefited. That is, if desert is not reason-giving in the case of deserved harms, is it also not reason-giving in the case of deserved benefits? An affirmative answer seems highly implausible; therefore, the authors must explain why the normative effect of desert is different in the two cases.

Third, just as NCJ1 makes a person’s desert a necessary condition of permissibly punishing him or defending oneself or others against him, so it also makes his deserving to be harmed a necessary condition of permissibly requiring him to compensate victims whom he has unjustifiably harmed. For if forcing a person to pay compensation is to intentionally cause him deprivation, and if (as the authors say) the infliction of “deprivation” is a harm, and if, finally, NCJ1 regulates all intentional inflictions of harm, then the principle must govern the requirement to pay compensation as a matter of corrective justice. It therefore seems incompatible with the liability rules of tort law—obviously the rules of strict liability but also the rules of fault liability, which require “fault in the act” but not culpability in the agent.

The authors, in correspondence, deny that imposing a duty of

compensation on a tortfeasor is a case of intentional harming; the harm is, rather, a side effect of ensuring that the victim is compensated. In a strict and literal sense, this is true: the harm itself is not intended, either as an end or as a means. But if what counts as an intended effect is understood in such a narrowly restricted way, the claim that there is a special constraint against intentional harming may be vitiated. For the same claim can be made about most inflictions of harm, not only in self-defense (“Yes, your honor, I harmed the police officer—yes, all right, *fatally*—but only as a side effect of preventing him from arresting me”) but in other cases as well (e.g., Jonathan Bennett’s case in which a pilot bombs a city with the intention of making its inhabitants *appear* to be dead, though this has as a side effect making them actually dead).<sup>20</sup> It seems that those (including Gardner, Tanguay-Renaud, and me) who accept the relevance of intention to permissibility must accept either a broader concept of an intended means or an understanding of the relevance of intention to permissibility that is different from that found in the Doctrine of Double Effect. I have discussed this problem elsewhere.<sup>21</sup> For present purposes, it is perhaps sufficient to note that if the authors claim that a court does not intentionally harm a tortfeasor when it requires him to pay a million dollars to someone he has harmed, they will have difficulty explaining how this is compatible with the claim that a criminal *does* intentionally harm a police officer whom he shoots in an effort to escape arrest.

Fourth, and perhaps most disturbing, by making desert a necessary condition of justified intentional harming, NCJ1 excludes both the possibility of a necessity or lesser evil justification for the intentional infliction of harm on an innocent person and the possibility that a person’s freely given consent could ever make it permissible to harm him intentionally. Despite the fact that this second restriction concerning consent has the appealing consequence of ruling out the permissibility of boxing, NCJ1 reduces the scope of possible justifications for intentional harming to an implausibly narrow range.

I should stress that these objections are based on a literal reading of NCJ1. But there is a passage in which the authors may seem to understand the principle differently. They write that

even when D goes beyond what E deserves in inflicting harm on E, the infliction may still be justified by other norms and reasons.

20. Jonathan Bennett, “Morality and Consequences,” in *The Tanner Lectures on Human Values*, vol. 2, ed. Stirling McMurrin (Salt Lake City: University of Utah Press, 1981), 110–11.

21. Jeff McMahan, “Revising the Doctrine of Double Effect,” *Journal of Applied Philosophy* 11 (1994): 201–12.

. . . This is the main role of NCJ1 . . . in the morality of punishment. It is not the case, as some suppose, that we are never justified in punishing those who do not deserve it, or never justified in punishing those who do deserve it more than they deserve. It is only the case that when we do punish these people we breach a duty to them and so, *ceteris paribus*, we owe them an apology as well as compensation. (130–31)

This seems incompatible with NCJ1, since moral justification entails permissibility, even if it does not entail that there is no violation of a duty. For this passage to be compatible with NCJ1, it seems that that principle would have to say, not that intentional harming is *permissible only if* the victim deserves it, but that it *breaches a duty if* the victim does not deserve it.<sup>22</sup> This uncertainty about the interpretation of NCJ1 emerges again, as we will see, in the authors' discussion of proportionality.

The various implausible implications follow from the apparent claim of NCJ1 that desert is a necessary condition of permissible intentional harming. Gardner and Tanguay-Renaud do gesture toward a view that could avoid these implications, preserve the relevance of desert to defense, and yet also avoid the implication that desert could justify harming those who pose a threat even in the absence of defensive effectiveness. According to this view, desert is not a necessary condition of permissible defense but does provide a positive reason to harm a person who poses a threat of wrongful harm, even when harming him would have no defensive effect. Yet desert alone can never be sufficient to justify the infliction of defensive harm; instead, the permissibility of defensive harming always depends on the presence of other reasons. This is because the reason for harming that is provided by desert is "systematically counteracted," or always offset, by a "reason of humanity" not to inflict harm on another person (125).

This too is a coherent view, but it is unclear why it might be thought to constitute a theoretical or substantive advance in relation to an account of self-defense based on a responsibility criterion of liability to defensive harm. Since it must deny that desert is a necessary condition of permissible harming, it avoids the implausible implications noted above and allows for the kind of justificatory plural-

22. In correspondence, the authors explain that they accept that "an action may be permitted under one permissive norm while remaining forbidden under a conflicting duty-imposing norm, and may be not permitted under one permissive norm yet still permitted under another." While I accept that an act can be *pro tanto* permissible under one principle and *pro tanto* impermissible under another, and that there can be cases of indeterminacy in which an act is neither permissible nor impermissible, I do not see how an act can be both permissible and impermissible all things considered. And what is at issue here is permissibility all things considered.

ism that the authors find appealing. It is therefore compatible with a responsibility criterion of liability. Indeed, in all cases in which this view implies that defensive harming is permissible, a responsibility criterion has the same implication, since desert entails culpability, and culpability entails responsibility. The view may therefore seem superfluous in a pluralist account that also comprises a responsibility criterion, unless there is reason to believe that moral responsibility for a threat of wrongful harm is sufficient for liability to defensive harm in some cases but not in others—that is, that in some cases, culpability and therefore desert are required in addition to moral responsibility.

I am open to this possibility. The authors, however, suggest a different way in which the view may be an important component of a pluralist account. They suggest that it could provide a better explanation of proportionality in defense than the one I offer. As they accurately note, my view is that moral responsibility is the basis of liability; that responsibility is a matter of degree, with culpability occupying the high end of the spectrum of responsibility; that liability, like responsibility, is a matter of degree; and that variations in the degree of liability are manifest in variations in the strength of the narrow proportionality requirement, which becomes less restrictive the higher the degree of a person's responsibility is. This means, for example, that it could be proportionate to inflict a certain harm on a culpable threatener to prevent him from inflicting a certain wrongful harm on oneself, but disproportionate to inflict that same defensive harm on an innocent threatener to prevent him from inflicting the same wrongful harm.

The authors argue that this way of accommodating the significance of culpability is inadequately explained. Why, they ask, should culpability have a role in the determination of proportionality when it has no role in the determination of whether defensive action is permissible at all? It would make more sense, they suggest, to adopt a pluralist account that offers one justification for self-defense against culpable threateners and a different justification for self-defense against innocent threateners, each with its own distinct proportionality restriction that deploys a metric appropriate to own criterion of liability.

This seems unnecessary. There is a plausible explanation of why, if responsibility is the criterion of liability, culpability would be relevant to narrow proportionality. This is that culpability is a *form* of responsibility. Culpability for a threat of wrongful harm is just moral responsibility for that threat in the absence of a fact-relative or evidence-relative justification or any excusing conditions.

But even if this is an acceptable answer to their initial challenge, they have a further, deeper objection to the claim that responsibility

(and therefore liability) diminishes when one has an excuse, with the degree of diminution proportional to the strength of the excuse. They argue that even if one can make sense of the idea that responsibility without culpability can vary in degree, the further claim that excuses diminish responsibility is doubtfully consistent with a proper understanding of the nature of excuse. Excuses, they note, “are available to one only qua responsible agent” (132). But “it is hard to reconcile this thought with the thesis that the more complete one’s excuse, the less one’s responsibility” (132).

It does seem right that responsible agency is a condition of any excuse. Lack of any capacity for responsible agency is not an excuse but an exemption from the demands of morality. A tiger that enjoys a missionary for its evening meal is not excused by virtue of its incapacity for responsible agency; rather, being beyond the reach of morality, it does not act impermissibly at all, so that there is nothing to be excused. (It is for this reason that I think Rodin’s discussion of “agency-defeating excuses” is unsatisfactory. A man who has been thrown down a well and thus threatens to crush a person at the bottom is not excused, since he has not *acted* at all; a fortiori he has not acted impermissibly, so there is nothing he can be excused *for*.) But the fact that excuse presupposes responsible agency does not subvert the claim that the degree to which one is responsible for an outcome diminishes in proportion to the strength of one’s excuse. For what is necessary for a person to have an excuse is that he should be capable of responsible agency. But it is not responsibility in the sense of the capacity for responsible agency that is diminished by an excuse. What is diminished is the extent to which the agent is responsible for a certain outcome. A person may, for example, have the highest possible capacity for responsible agency and yet bear only minimal responsibility for some harm he has caused by virtue of being nonculpably ignorant that his action involved certain risks. Hence, I think the authors are correct when they qualify their remarks about excuses and responsibility by conceding that “perhaps there is some equivocation . . . here about the relevant sense of ‘responsibility’” (132).

I will conclude with two brief comments on the authors’ suggestion that a harm’s being *undeserved* constitutes a positive reason not to inflict it. As they mention in a footnote, I have expressed the concern that if what they call “undesert” is a reason not to inflict an undeserved harm, the question arises whether it is also a reason not to bestow an undeserved benefit. They respond by remarking, of the bestowal of benefits, that “it’s one thing to rule out its permissibility and another to recognize, as we do, that there’s a reason against it” (127 n. 26). But rather than defending the claim that the reason against it is undesert itself, they cite “the cost to the generous person.”



This, however, seems irrelevant, since their claim about undeserved harm is that the positive reason not to inflict it is the undesert itself, not that it is sometimes costly to an agent to inflict an undeserved harm. I suspect, though, that they need not worry about undeserved benefits, for absence of desert is not in fact a reason against the infliction of harm either. The belief that it is may derive from taking too seriously what is merely a manner of speaking. In the case of punishment, the claim that “he does not deserve it” is a forceful objection because it is widely assumed that desert is a necessary condition of justified punishment (and perhaps even a conceptual condition of punishment itself: in the days of ordinary language philosophy, it was sometimes claimed that punishment of the innocent is a conceptual impossibility). Thus, to say of a punishment that “he does not deserve it” is to say that a condition that is *essential* to the justification of punishment is missing. But in other instances of the intentional infliction of harm, such as self-defense, the claim that “he does not deserve it” is primarily rhetorical. Taken literally, it merely asserts the absence or inapplicability of one of a number of possible forms of justification for harming—in the same way that “he didn’t ask for *that*” might literally refer to the absence of consent but is generally asserted only rhetorically. If one were to say, of a person ordered to pay significant compensation to the victim of an accident caused by minor carelessness, that he does not deserve such a heavy burden, the proper response would be, “Of course not. The infliction of deserved harms is the business of criminal law, not the law of torts.”