Individual Liability in War: A Response to Fabre, Leveringhaus and Tadros

JEFF McMAHAN
Rutgers University

This article is a response to commentaries on my book, *Killing in War*, by Cécile Fabre, Alex Leveringhaus and Victor Tadros. It discusses the implications of the approach I have defended for the morality of war for such issues as internecine killing in war, humanitarian intervention and the bases of individual liability to attack in war.

I have sought to develop and defend an account of justified killing in war that assumes that war is morally continuous with lesser forms of violent conflict and that the morality of individual action in war is therefore governed by the same principles that govern individual self- and other-defence. I am both honoured and grateful that Cécile Fabre, Alex Leveringhaus and Victor Tadros have thought it worthwhile to consider my ideas and arguments as carefully and thoroughly as they have in their commentaries. Readers may feel cheated, however, for rather than tear down the edifice I have tried to erect, all three commentators have either explored the implications of the type of account I have defended for issues that I have not addressed or else suggested different ways of understanding the account with the aim of identifying the best version. By contrast with those who may be disappointed at the comparative lack of philosophical gore in the preceding essays, I am profoundly relieved. In what follows, I will indicate points of agreement and disagreement and seek to provide further elucidation of the issues on which the commentators have focused: internecine killings in war, humanitarian intervention, and the basis and scope of an individual's liability to incur harm in war. In preparing this response, I have learned a great deal from all three commentators.

FABRE

Following her general practice of discussing issues in the morality of war that have hitherto been largely unexamined, Cécile Fabre considers the permissibility of self-defensive killing by soldiers of their own officers, when the latter threaten to kill them if they disobey an order. She distinguishes between cases in which what an officer orders a

Utilitas Vol. 24, No. 2, June 2012
doi:10.1017/S0953820812000131

soldier to do is morally impermissible and cases in which the soldier has a moral duty to do what he is ordered to do, perhaps even independently of his having been ordered to do it. She concludes that the self-defensive killing of an officer is permissible when the order he gives is unjust and may also be permissible even when the order is just. I find her arguments generally persuasive in cases involving unjust orders. But I am sceptical of her view about cases involving just orders and will thus devote most of my remarks to these latter cases.

It may be that Fabre's permissive view of internecine killing in war derives in part from her claim that O's threat to kill S constitutes 'a lethal attack' against S. For it is plausible to suppose that S has a right of self-defence against a lethal attack prompted by nothing more than his refusal to obey an order. Yet I think that Fabre ought not to treat O's threat to kill S as a lethal attack. Nor should she say that when O threatens to kill S if he refuses to obey, O 'violate[s] S's right not to be killed'. For a threat of lethal attack, whether or not it is accompanied by a declared threat *to* attack, is quite different from an actual attack. I recognize, of course, that unless a declared threat is a bluff, both a threat of attack and an actual attack raise the probability that the victim will be killed. But the difference, in the cases that Fabre discusses, is that the threatened attack is conditional on S's disobedience. There will be no actual attack if S obeys. Thus, if S does in fact obey, so that O does nothing more than threaten him, O will not have violated S's right not to be killed, though he may have violated S's right not to be wrongly threatened, or not to be put at significant risk of being killed.

Fabre's apparent conflation of threats of attack with actual attacks can be explained if we attend to a further distinction. When she says that threats are attacks, she seems to have in mind cases in which the victim will have no option of defence if he defies the threat. She writes, for example, 'that by placing a gun to S's head, O is attacking him'. Since O literally has a gun to S's head, S will have no option of self-defence if he defies O's threat. He must either kill O now, or obey the order, or submit to being killed. But often when a person is threatened with death for refusal to obey a command, there are two options for defensive action: (1) immediate action – that is, after the threat has been made but when the defender has neither obeyed nor disobeyed the command; and (2) action after the defender has defied the threat and the threatener is about to attack, or is actually attacking. Fabre's remarks are intended to apply only to cases, such as those involving a gun to the head, in which there is no option of defence subsequent to S's defying the threat. But not all cases are of this sort and there is no reason to limit the discussion of internecine killings to these cases only.

What if S has the option of defence either before or after disobeying O's unjust order? I will assume that it is permissible for S to refuse to

obey an unjust order, even when the wrong he is ordered to commit is comparatively trivial. It may nevertheless be impermissible for S to kill O without first disobeying, for before S disobeys, O threatens him only conditionally. S would not, for example, be permitted to kill O without first disobeying if the probability of successful defence would not be much diminished if S were to engage in defensive action only after disobeying, when O would be unconditionally committed to killing him. But if S refuses to obey and O then attempts to fulfil his threat, S is clearly justified in killing O in self-defence. For O is at this point morally responsible for an unjust attack on S's life and thus has made himself liable to be killed. It makes no difference now whether what O ordered S to do was seriously or only slightly wrong. If, however, the wrong that S has been ordered to commit is serious and the probability of S's being able to defend himself effectively would be more than minimally diminished after he had disobeyed, then S may kill O immediately in self-defence, without first disobeying. For in the circumstances S is not required for O's sake to accept an increased risk of being wrongfully killed by O.

Next consider cases in which S would be morally justified in doing what O orders him to do. Suppose that S would be justified even in the absence of the order and that, with the order, it becomes his duty to do as ordered.

There are questions to which the answers are obvious in cases involving unjust orders but not in cases involving just orders. Two such questions are these. (1) Is O permitted to threaten to kill S as a means of compelling him to obey? (2) Is O permitted to kill S if S refuses to yield to the threat and continues to disobey the just order? I will also distinguish and separately address two further questions. These are: (3) Is S permitted to kill O in self-defence as an immediate response to O's threat? (4) Is S permitted to kill O in self-defence after disobeying O's order, when O attempts to fulfil his threat to kill S?

First, is it permissible for O to threaten to kill S for disobeying O's just order? Except perhaps in certain highly unusual circumstances, it is permissible for O to threaten to kill S *if* it would be permissible for O actually to kill S if the latter were to continue to refuse to comply. In general, the permissibility of *doing* X in conditions Y implies the permissibility of *threatening* to do X if Y – though it does not follow that the permissibility of threatening to do X if Y entails the permissibility of actually doing X if Y. Thus, for there to be a serious question whether O may threaten to kill S, we must assume, for the moment, that it would be impermissible for O to kill S.

Even so, it seems obvious that it would be permissible for O to threaten to kill S if the consequences of S's disobedience would be very bad, the threat would have a high probability of success *and* the threat

was a bluff – that is, if O would not in fact kill S if S were to disobey. But Fabre says at the outset that she is not concerned with empty threats and we can follow her in putting them aside.

That leaves two possibilities: either O conditionally intends to kill S or is uncertain what he will do if S refuses to comply. While it may be impermissible in itself for O to intend conditionally to do what it would be impermissible for him to do, this is highly controversial. There is, however, one difference between conditionally intending and remaining undecided that is uncontroversially relevant, which is that a conditional intention would nearly always raise the probability that S will be killed by more than being undecided would. And on the assumption that it would be impermissible for O to kill S, it is clearly relevant to whether it is permissible for O to threaten to kill him what the probability is that he will fulfil his threat if S continues to disobey. Other considerations that are relevant include two already mentioned: the expected consequences of S's failure to obey the just order and the probability that the threat will induce S to obey. These three considerations together determine whether, if the threat would otherwise be permissible, it would be proportionate. And there is also a necessity condition: roughly, that no less risky an alternative, such as a threat of lesser harm, would have an equally high probability of inducing compliance.

Assuming that S has a moral duty to obey and that this duty is owed not just to O but to those who will be saved if he fulfils it, he is liable, it seems, to some level of coercion in order to secure his compliance. Given that it is possible that a threat to kill him would, in the circumstances, be both necessary and proportionate in relation to the good that would be achieved through his obedience, it seems that he could be liable to be threatened with death, even when the increase in the probability of his being killed is taken into account, and even if it would be impermissible for O actually to kill him if he still refuses to obey. For it could be permissible for O to *risk* acting impermissibly as a means of coercing S to do his duty. It also seems relevant that it is in S's power to avoid the risk of being killed simply by doing his duty.[1]

Despite the fact that it might in principle be permissible for O sincerely to threaten to kill S, it is unlikely in practice that an officer can be justified in threatening to kill a soldier under his command as a means of securing obedience to a just order. This is so for a variety of reasons: the officer may have other means of achieving what he has ordered the soldier to do (he might, for example, do it himself); he might induce the soldier to obey by other means; such threats may undermine

[1] For an extended defence of the relevance of this consideration, see V. Tadros, *The Ends of Harm: The Moral Foundations of Criminal Law* (Oxford, 2011).

the cohesive relations between the officer and those under his command on which his *de facto* authority rests, and so on. But, as I noted earlier, the threat is likely to be wrong only if it would be wrong to fulfil it. Thus far I have been assuming that it would be wrong for O to kill S. Are there any reasons to doubt this? Might there be circumstances in which it would be permissible for O to kill S for disobedience? If so, it would almost certainly be permissible for O to threaten to kill S in those circumstances.

I will assume that O's killing S could not be justified as a matter of punishment. That is not just because O lacks the authority to punish S but also because, assuming that S deserves punishment, killing him would be disproportionate in relation to his desert. But might killing S be justified as a means of saving those innocent people who might otherwise be saved by S if he were to obey O's order? As Fabre points out, it would, in one respect at least, be counterproductive to kill S, for what O wants S to do is to save the innocent people, which he cannot do if he is dead. (If a *threat* to kill S constitutes a 'lethal attack', as Fabre suggests, then a lethal attack might indeed cause S to save the innocent people rather than preventing him from being able to save them.) But, as she also recognizes, there are causal paths by which killing S might result in the saving of the others. By killing S, for example, O might intimidate other recalcitrant soldiers in his unit, thereby ensuring that one of them, S1, will do what S ought to have done. If there is a justification for killing S, therefore, it must be that this is the only means, or perhaps the best means, of ensuring that *someone* will obey O's just order. (Note that if O has this reason to kill S, which is neither punitive nor defensive, this enhances the credibility of his threat to kill S, thereby making it more likely that S will obey and hence less likely that O will actually fulfil the threat.)

Fabre rightly observes that to kill S to induce S1 to obey O's order is to use the killing of S, and therefore S himself, as a means of achieving O's just aim. And she is right to say that I follow Warren Quinn in thinking that there is a special constraint against harmfully using a person in this opportunistic way, one that is especially strong when the victim is in no way responsible for a wrong or a threat of wrongful harm.[2] But unless some form of moral absolutism is true, it does not follow that opportunistic killing cannot be permissible. If, for example, there are enough innocent people on the track who will otherwise be killed by a runaway trolley, there can be a lesser evil justification for pushing a large man onto the track to stop it. More importantly for our purposes,

[2] For the distinction between opportunistic and eliminative forms of intentional harming, see Warren S. Quinn, 'Actions, Intentions, and Consequences: The Doctrine of Double Effect', *Philosophy and Public Affairs* 18 (1989), pp. 334–51, at 344.

people can act in ways that make them liable to be opportunistically killed. Suppose, for example, that yesterday I tampered with the brakes in your car in the hope of killing you. Now your car is careering out of control and the only way you can prevent it from going over a cliff is to steer it into me.[3] In running your car into me, you would be using me, and in a way you know would cause my death, as a means of bringing your car to a stop. But it seems that in doing so you would not be wronging me or violating my right not to be opportunistically killed. For by tampering with your brakes I made myself liable, in the circumstances, to be killed in this way.

It does seem to me that the threshold for liability to be killed opportunistically is higher than the threshold for liability to be killed defensively or, in Quinn's more expansive term, 'eliminatively'.[4] It may well be, for example, that while non-culpable responsibility for a threat of wrongful harm can be sufficient for liability to defensive killing, culpability is necessary for a person to be liable to opportunistic killing. But it may also be that S is highly culpable for refusing to obey O's just order and that unless O kills him opportunistically as a means of coercing S1, many innocent people will be killed by unjust combatants. In such a case, it may be permissible for O to kill S, not on the basis of a lesser evil justification but because S, through his own culpable wrongdoing, has made himself liable to be opportunistically killed. If, by contrast, there are reasons why S is excused for refusing to obey, so that his action is not culpable, we should probably conclude that he is not liable to opportunistic killing. Much depends on the facts of the particular case.

It is true, as Fabre points out, that S's wrongdoing involves *allowing* harm to occur rather than *doing* harm himself. And she is again right to suppose that I think this makes a moral difference. I accept that, if all other relevant considerations are equal, S might be liable to be opportunistically killed to prevent him from doing a harm (as in the case in which you run your car into me to prevent my action from killing you) but *not* liable to be opportunistically killed to prevent an equivalent harm that he would otherwise wrongly allow to occur. But the moral asymmetry between doing and allowing is nicely counterbalanced in this case by the fact that S has at least two special reasons not to allow those harms to occur that he has been ordered to prevent. The more important of these is that he has a professional, role-based duty as a soldier to prevent unjust combatants from harming innocent people. The other is that he is specially related to those who would be harmed:

[3] I borrow this example from Jeff McMahan and Robert McKim, 'The Just War and the Gulf War', *Canadian Journal of Philosophy* 23 (1993), pp. 501–41, at 514.

[4] Quinn, 'Actions, Intentions, and Consequences', p. 344.

they are his fellow citizens, some of whom are also his comrades-in-arms. These considerations may make his allowing these people to be killed morally similar to killing them. As an analogy, consider that many people, including many philosophers, have thought that parents who allow their child to starve to death have killed, or murdered, their child.[5] These people have mistakenly allowed their moral reaction to the parents' action to influence their classification of the parents' mode of agency. For unless they have actively prevented the child from having food, the parents have not killed their child but have allowed it to die, just as anyone else who could have fed it but did not would have done. Yet we judge the parents' action to be the moral equivalent of murder because of their special relation to the child and their role-based duty as its parents to care for it.

In summary, whether O is permitted to fulfil his threat to kill S depends crucially on whether the killing is necessary to ensure that someone else will obey O's just order. It also depends, of course, on whether S's refusal to obey is sufficiently culpable to make him liable to be killed as a means of inducing someone else to do what he ought to have done. Or, if S is not liable (perhaps because he has an excuse that excludes culpability), it depends on whether what O has ordered is sufficiently important that there is a lesser evil justification for killing S as a means of achieving it. As I indicated before, much depends on what the non-moral facts are.

Finally, is S permitted to kill O in self-defence in cases in which O's order is just? The answer depends on whether it is permissible for O to kill S. Suppose first that it is permissible for O to kill S because conditions are such that S's action has made him liable to opportunistic killing. In these circumstances, it is not permissible for S to kill O in self-defence, either before or after disobeying O's order. The reason is that, because the only bad effect of O's action will be to inflict harm on S to which he is fully liable, O cannot be liable to defensive action by S; nor is there any other form of justification, such as one based on consequences, for S's killing O in these circumstances. (It is possible that it would be permissible for S to kill O in self-defence if O's justification for killing S were a lesser evil justification rather than a liability-based justification. I will not pursue this complication here.)

But suppose that S's disobedience is insufficient to make him liable to opportunistic killing and that there is no necessity justification for killing him, so that it is not permissible for O to kill S. Is it then permissible for S to kill O to prevent himself from being wrongly killed?

[5] For an early example, see Philippa Foot, 'Abortion and the Doctrine of Double Effect', originally published in 1967, reprinted in her *Virtues and Vices* (Oxford, 1978), pp. 19–32, at 26.

Given that there may be actual cases of this sort in which S will have the option of self-defence even after refusing to obey, we should distinguish two questions and address them both. (1) Is it permissible for S to kill O in response to the threat alone, without first disobeying? (2) Is it permissible for S to kill O after disobeying, when O will otherwise kill him?

The answer to the first question is that killing O is not permissible, for in the circumstances it is neither a necessary nor a proportionate defensive option. At that point, S has two options for preserving his life. One involves killing O and allowing a number of innocent people to be killed. The other, which is to obey O's order, spares O's life and saves the same innocent people. Even if the second option involves greater risks or costs to S, this is irrelevant because, by hypothesis, this option is already S's *duty*. This is what he must do. He may not kill O to enable himself to avoid doing his duty. It seems, in fact, that it is irrelevant to the permissibility of S's killing O at this point that the threatened attack against which S would be defending himself would be wrong.

The second of these two questions is quite different. Once S has disobeyed, he no longer has any option for self-preservation other than killing O. Even so, it may seem that it remains impermissible for him to kill O. For it was his own wrongful action in disobeying that produced the conditions in which killing O is now his only option for self-preservation. I think, however, that because O's attack on S – for it is now more than merely a threat – is not justified by S's wrongful disobedience, O makes himself liable to defensive killing by S. That O's killing S would be unjustified indicates, in the circumstances, that S does not satisfy the conditions for liability to opportunistic killing. There are two possible explanations of why this is so. One is that killing him would be disproportionate in the narrow sense in relation to the expected good that might thereby be achieved (perhaps because it would have only a relatively low probability of saving the innocent people). The other is that S is not culpable for refusing to obey. But if either or both of these explanations are true, it seems intuitively plausible that S is permitted to kill O in self-defence, even though his refusal to obey, which triggered O's attack, was wrong.

LEVERINGHAUS

Leveringhaus refers to the type of approach to the morality of war that I defend in *Killing in War* as the Neo-Classical View and to the more traditional approach to which it is opposed as the Legalist View. These are accurate descriptive labels and I will follow him in using them here. He aims in his commentary to explore the implications of the Neo-Classical View for the morality of humanitarian intervention

and to compare them with those of the Legalist View, in an effort to determine which of the two approaches provides the more plausible account.

Advocates of the Legalist View have found it difficult to recognize the permissibility of humanitarian intervention. This is not necessarily an embarrassment, as there are serious and quite general moral objections to humanitarian intervention. But the Legalist View's objections are not among them; they are instead rather bizarre and help to expose the deeper implausibility of some of the theory's foundational assumptions, such as that the principal agents in war are states and that relations among states in war are governed by the same moral principles that govern relations of aggression and defence among individual persons. These assumptions constitute what Michael Walzer, the foremost contemporary exponent of the Legalist View, calls the 'domestic analogy'. According to the domestic analogy, states are analogues of persons and have rights that are analogous to those of persons. Assuming, as Mill's 'harm principle' asserts, that individual persons have a right against coercive interference in their lives when their action causes no harm to others, it follows from the domestic analogy that states have a right against forcible interference in their internal or domestic affairs, provided that their acts do not cause harm to other states. The persecution of one group by another within the same state is thus analogous to an individual's harming himself. It is the state harming itself. Forcible intervention by another state to stop the persecution is therefore an instance of paternalism. And just as paternalism in the case of an individual person violates that person's right of autonomy, so paternalism in the case of a state violates that state's right of political sovereignty in a matter that is wholly self-regarding.

This is obviously not a plausible way of conceiving of domestic persecution or of external intervention to end it. And adherents of the Legalist View have usually conceded that humanitarian intervention can be permissible to stop large-scale violations of basic human rights – for example, in cases of genocide, mass expulsion, or enslavement. But these concessions seem ad hoc; they have no principled basis within the Legalist View.

In contrast to the statist character of the Legalist View, the Neo-Classical View is individualist, in that it holds that the principal agents in war are persons, not states, and that the justifications for acts of war by individuals must be of the same kinds that justify acts of harming or killing by individuals outside the context of war. Even in war, morality continues to apply to individual persons, not to analogues of persons.

According to the Neo-Classical View, the strongest justification for killing in war appeals to the claim that those who are killed have made

themselves morally liable to be killed as a means of preventing (or, rarely, correcting) wrongful harms for which they are to some degree morally responsible. That is, they have forfeited or lost the right they would otherwise have not to be killed as a means of preventing this harm. There is then a just cause for war when the people it is causally necessary to attack as a means of preventing the wrongful harms are both liable to attack and sufficiently numerous that the attacks require military means.

Whereas the Legalist View insists that preventing the violation of human rights in another state can be a just cause for war only when the violations would otherwise be massive in scale, the Neo-Classical View claims that even small-scale violations can justify humanitarian intervention, provided that military action is necessary and proportionate.[6] Suppose, for example, that a tyrannical regime has arrested and begun to torture 100 political opponents who have done nothing other than participate in non-violent protests against the regime's oppressive policies. Releasing those prisoners could be a just cause for a highly limited war – for example, a military incursion by commando forces of another state that will require the killing of military guards, torturers and other official personnel at the prison as a means of extracting the prisoners and getting them to safety.

One might object that so limited an action would not constitute a *war*. It seems undeniable, however, that it would be an *act of war*, albeit one on a sufficiently small scale that it might be unlikely to provoke retaliation and escalation (and would presumably be proportionate only if it would not). That it would be an act of war in *legal* terms is revealed by the fact that it would activate the law of war, so that the commandos, being military agents of a state engaged in an attack against armed agents of another state, would have combatant status and hence, if captured, be exempt from punishment and entitled to the rights of prisoners of war. If the incursion were not to provoke retaliation, it would simply be a limited, unilateral war. That wars can be unilateral is shown by the fact that a nuclear first strike that would completely destroy an adversary's retaliatory capacity would certainly count as a nuclear war. And if the incursion were to prompt retaliation, the larger bilateral war would be dated as having begun with the incursion, not with the retaliation.

[6] In an earlier article, I noted that, assuming that war involves killing (a contingent rather than necessary truth), 'only aims that are sufficiently serious and significant to justify killing can be just causes. Beyond this, however, considerations of scale are irrelevant to just cause.' See 'Just Cause for War', *Ethics and International Affairs* 19 (2005), pp. 1–21, at 11. I have since learned from recent conversations with Kieran Oberman that he has independently arrived at the same conclusion: that even a small-scale violation of human rights can constitute a just cause for a proportionally small-scale war.

Some may continue to insist that the incursion cannot count as a war because of its limited scale. But this is unimportant, for nothing substantive, at least where morality is concerned, depends on what word we use to describe the relevant events. We can simply say that in this case there is a just cause for a limited military incursion that involves killing armed agents of another state.

Leveringhaus's discussion, though, focuses less on the permissibility of humanitarian intervention than on the permissibility of defence against it. He makes the important but little noticed observation that, according to the Legalist View, while it is not permissible for members of a state's military to commit atrocities against their own population, it *is* permissible for them to defend themselves against those, whom I will call *just interveners*, who participate in a justified humanitarian intervention intended to stop an atrocity. This is because just interveners are combatants and the Legalist View holds that all combatants are legitimate targets for enemy combatants at any time during a state of war. The reason usually given to explain why they are legitimate targets is that they pose a threat to others and hence are not innocent in the sense deemed relevant by the Legalist View. Because they pose a threat, they are liable to defensive attack. In this the Legalist View follows the law on which it is modelled. The law of armed conflict does not prohibit soldiers engaged in the commission of an atrocity from conducting a defensive counterattack against enemy combatants who attack them in an attempt to stop the atrocity. For although atrocities are of course illegal, defensive attacks against enemy combatants are not, even if they enable those who conduct them to commit an atrocity.

But while the law permits soldiers to defend themselves even while they are engaged in the commission of atrocities, Leveringhaus rightly sees that morality offers no such permission. If it is impermissible for soldiers to torture and kill their innocent fellow citizens, even under orders from their government, and if their committing such an atrocity is what morally justifies military intervention against their state, it is absurd, and even doubtfully coherent, to suppose that it could be morally permissible for them to enable themselves to continue by killing those who justifiably try to stop them. This is, indeed, the clearest instance I know of in which the law of armed conflict permits what morality prohibits.

Leveringhaus's major concern, however, is not with self-defence by the perpetrators of atrocities but with the permissibility of defence against just interveners by soldiers of the target state who are *not* among the perpetrators. I will refer to such soldiers as 'non-participating defenders'. Leveringhaus notes that, according to the Legalist View, these soldiers 'can potentially be viewed as moral equals.

This is so because the target state's campaign against the intervening state resembles a more traditional self-defensive war. He is right that the Legalist View recognizes no difference in moral status between just interveners and non-participating defenders, but his explanation is overly subtle. The correct explanation lies again in the same crude doctrine that licenses defensive killing by the perpetrators of atrocities – namely, that all combatants may permissibly be killed by enemy combatants during a state of war.

What about the implications of the Neo-Classical View? Leveringhaus says that although 'the Neo-Classical View is closer to our intuitions about [humanitarian intervention] than the Legalist View . . . when it comes to encounters between intervening combatants and perpetrators of Atrocity Crimes, the application of the Neo-Classical View' to the killing of just interveners by non-participating defenders 'remains problematic'. It is problematic, he suggests, because, 'unlike perpetrators of Atrocity Crimes, [non-participating defenders] are the victims of (just) aggression without having posed an unjust threat'. He therefore seeks to draw out the implications of the Neo-Classical View by advancing an 'Argument from Derivative Liability', based on an appeal to ideas borrowed from the law of complicity. It is unclear, however, how the logic of this argument is supposed to work. His argument that non-participating defenders are complicit in the wrongdoing of the perpetrators of atrocities seems clearly correct. The question is what this shows. I think he succeeds in showing that non-participating defenders are liable to attack. But that is uncontroversial. They are liable according to the Legalist View because they are combatants. And they are liable to defensive attack according to the Neo-Classical View because they will otherwise cause wrongful harm to just interveners without moral justification. They lack justification because the just interveners have neither consented to be attacked nor made themselves liable to be attacked, and there is clearly no lesser evil justification for attacking them. But if the rights of just interveners not to be attacked have been neither waived, nor forfeited, nor overridden, it seems that attacking just interveners must be morally unjustified. Even a Hobbesian account of self-defence, which asserts that even the most egregious wrongdoers have a right of nature to defend themselves against their victims, does not imply that non-participating defenders have a *moral* justification for killing just interveners.

Everyone can agree, therefore, that non-participating defenders are liable to attack. But what Leveringhaus claims is implied by the Neo-Classical View is only that non-participating defenders are not permitted to fight in defence against just interveners. Yet this does not follow from the Neo-Classical View's implication that they are liable to attack. For the Neo-Classical View accepts that it is possible

to become liable to attack by acting permissibly – for example, when one permissibly acts in a way that exposes innocent people to a tiny risk of significant harm that, through bad luck alone, eventuates in an imminent threat of significant harm. So the fact that non-participating defenders are liable to attack does not by itself entail that their defence of others is impermissible. Leveringhaus seems to suggest that it is instead the fact that their defensive action makes them complicit in the commission of an atrocity that entails that this defensive action is impermissible. That is, his claim might be that any action is impermissible if it makes the agent complicit in an atrocity. This may be true, but there is a related yet simpler explanation of why the Neo-Classical View implies that it is impermissible for non-participating defenders to engage in defensive military action against just interveners. This is that their attacking just interveners involves intentionally harming people who are not liable to be harmed, and doing so as a means of achieving an unjust aim – that is, to shield the perpetrators of an atrocity.

The impermissibility of defensive action by non-participating defenders may even be implied by the impermissibility of defence by the *perpetrators* of atrocities. Most people think that *self*-defence is in general more strongly justified than third party defence of others, or '*other*-defence'. They believe, in other words, that there is an *agent-relative* dimension to the permissibility of defensive action. As a general matter, I think that is wrong. While there may be some cases in which part of the justification for self-defence is agent-relative, I think that in most cases the considerations that justify self-defence also justify other-defence. The justification in both cases is agent-neutral. But I am in the minority on this issue. Most people seem to believe that there are many cases, especially ones involving morally innocent threateners, in which an individual is justified in defending himself but third parties, and especially disinterested third parties, are not permitted to intervene on his behalf. Yet few people believe that there are cases in which, even though an individual has no right of self-defence (for example, because he is a wrongful aggressor), it is nevertheless permissible for third parties to defend him. Admittedly, a few people, such as Augustine, have held that because altruism is a higher motive than self-interest, it can be permissible to kill in defence of another person though not in defence of oneself. But this view is uncommon, because it applies not just to defence against innocent threateners but also to self-defence against culpable threateners.[7] A somewhat more plausible view is that certain

[7] Augustine's view is actually more complicated than many people suppose. His doubts about the permissibility of self-defence are confined to individual or private self-defence and do not apply to collective self-defence through war, which he treats as an instance of

special relations can ground special permissions to defend others. For example, even when a wrongdoer is not permitted to defend himself, it might nevertheless be permissible for his parent to defend him. But even if this view is correct, it is of little relevance here, for only the most important special relations could have this moral power, and neither citizenship of the same state nor membership in the same military organization is among them.

Thus, although my reasons are in some instances slightly different from his, I agree with Leveringhaus that the Neo-Classical View implies that it is impermissible for the perpetrators of atrocities and for non-participating defenders to fight defensively against just interveners. Surrender seems to be the only permissible option they have, though perhaps they might also be permitted simply to flee. The failure of the Legalist View to recognize the moral impermissibility of self-defence by the perpetrators of atrocities, and of other-defence by non-participating defenders, helps to reveal the implausibility of its general claim that unjust combatants always act permissibly when their attacks are directed towards, and affect, only enemy combatants. For there is no general objective moral difference between a recognized atrocity – for example, the murder of civilians – and an unjust war of aggression. Indeed, unjust war often is, in moral terms, a greater atrocity than most instances of recognized atrocities. The Nazis, for example, committed a great many recognized atrocities. Many of these involved the murder of young male civilians. But Nazi soldiers also killed a very great many young male soldiers of the many countries they invaded. Is it really plausible to suppose that there is a significant moral difference between killing a young man when he is a civilian and killing him after morally compelling him to become a soldier in order to defend himself and other innocent people? Is there really any difference between breaking into a man's house and killing him before he can try to defend himself and breaking into his house and killing him as he tries to defend himself? The difference between a soldier's committing an atrocity and his killing just combatants in an unjust war often amounts to little more than that.

There are, obviously, a great many mitigating conditions that apply to combatants who kill just combatants in an unjust war that do not apply to those who commit recognized atrocities. These are of course relevant to issues of culpability, blame and liability. But the issue that Leveringhaus is concerned with is the *permissibility* of killing. It is on this issue that the Legalist View and the Neo-Classical View diverge.

other-defence. He also holds that individual self-defence is impermissible only according to divine or natural law, not according to human law, which he concedes must permit it. I am indebted to Henrik Syse for clarification of Augustine's views.

The former claims that while soldiers are not permitted to commit atrocities, they are permitted to kill enemy combatants as a means of achieving goals that are unjust. The latter denies that they are permitted to do either. My claim is that because an unjust war may involve even more extensive violations of basic human rights than a recognized atrocity, the position taken by the Legalist View is arbitrary and implausible.

TADROS

Victor Tadros argues that there is 'a deficiency in McMahan's general account of liability', which is that the account fails to accommodate instances of liability based on the absence of a right not to be harmed that is correlative with the presence of an enforceable duty to bear some harm. The absence of the right need not, he notes, be the result of the person's having lost or forfeited it through action for which she was responsible. She may instead have acquired the duty to incur the harm in some other way.

This is an important point, though I interpret it and its significance differently from the way Tadros does. The cases that he cites in which it is permissible to harm a person because she has an enforceable duty to incur that harm – taxation, conscription, drawing the short straw, and so on – are not instances of liability to be harmed, as I understand the notion of liability. As Tadros recognizes, it is essential to liability as I understand it that it involves the *loss* of a right, usually through action that makes a person responsible for a wrong, not merely the *absence* of a right. A foetus, for example, may lack a right not to be killed. But it does not follow that it is morally liable to be killed, in the sense in which I use the term 'liable'. Neither, of course, does the foetus have an enforceable duty to die or to allow itself to be killed. There are reasons why an individual might lack a right not to be harmed other than that the individual has forfeited it or has an enforceable duty to bear the harm.

What is important about Tadros's cases is that they reveal a form of justification for harming people that is often overlooked. There are three familiar forms of justification for harming and killing that I discuss in *Killing in War*. One is what I call a liability-based or, more simply, a liability justification. In the vocabulary of rights theory, a person becomes liable to be harmed in a certain way when he has *forfeited* his right not to be harmed in that way. There are also cases in which a person's consenting to be harmed can justify harming him, or contribute to the justification for harming him. In these cases, rights theorists say, he has *waived* his right not to be harmed. The third familiar form of justification may apply when the victim of a harm has neither waived nor forfeited his right not to be harmed. In such

cases, harming the person may be justified as the 'lesser evil' if the consequences of not harming him would be far worse from an impartial point of view than the consequences of harming him. In these cases rights theorists say that he retains his right not to be harmed but that it is *overridden*. There is, as I indicated in the previous section, a possible fourth form of justification that some writers believe has an essential role in the justification of self-defence – namely, an 'agent-relative' permission. This is a permission to harm another person either as a means or a side effect of avoiding a harm to oneself in certain circumstances in which none of the other justifications applies. It is sometimes claimed, for example, that it can be permissible for one person, in saving his own life, to kill another, either intentionally or as a side effect, even when the other has neither waived nor forfeited his right not to be killed. Since exactly one person will die either way, the justification cannot be a lesser evil justification (unless, perhaps, the other person would otherwise inevitably die very soon from a different cause). The claim is that the justification is instead that each person is permitted to give a certain degree of priority to his own interests over those of others, at least in certain conditions. It is less clear how to articulate this form of justification in the language of rights. One suggestion is that this is a form of *necessity* justification, with self-interest rather than an impartial assessment of the consequences as the source of the 'necessity'.[8] According to this interpretation, the victim's right is overridden, though by consequences assessed from an agent-relative rather than an agent-neutral perspective.

As I understand him, Tadros is claiming, rightly, that there is a fifth form of justification for harming: that it can be permissible to inflict a certain harm on a person when that person has an enforceable duty to suffer that harm. The person may have neither waived nor forfeited a right not to suffer that harm; nor is it the case that he retains that right but it is overridden; rather, he lacks the right because he has a duty to suffer the harm. Tadros says of such a person that he is liable to be harmed. I withhold that description, reserving it instead for a more limited class of cases in which the reason it is permissible to harm the person is that he has forfeited his right not to suffer that harm. This seems to me a comparatively unimportant disagreement between us – a disagreement about whether the word 'liability' has a more or less expansive scope. Tadros suggests that it is a reason to favour his way of using the word that it coincides, at least in some respects, with the way it is used in the law. I am aware that in certain areas of the law 'liability' refers to a person's vulnerability to the imposition of a

[8] I owe this suggestion to Uwe Steinhoff.

duty. But I am unmoved by this; when I began to use the term in discussions of self-defence and war I was conscious of using it in a way that was related to but distinct from the way it is used in law. It seems to me, moreover, that there is a good reason not to expand it in the way Tadros recommends. This is that a person's having forfeited his right not to have to suffer a certain harm and his having an enforceable duty to suffer that harm are quite distinct moral phenomena that offer different justifications for harming that person. While liability in my sense to suffer a particular harm generally (though not always) entails an enforceable duty to submit to that harm, an enforceable duty to suffer a harm generally does not entail liability to suffer it. Given that these are conceptually and morally distinct notions, clarity is best served by assigning them different labels.

The practical significance of this fifth form of justification for harming a person, considered entirely on its own, seems quite limited. Most of the examples that Tadros cites are ones in which the enforceable duty a person has to incur a harm is one that he can fulfil himself: the wealthy Briton submits her own tax payment, the conscript presents himself for duty and goes to war without resistance, the saviour of humanity gives up his life and Bertha remains behind to be devoured by the tiger. In each case, if the person with the duty to incur a harm will fulfil the duty without coercion, he or she retains a right not to have the harm imposed by a third party. (Tadros concedes a closely related point in his discussion of consent.) Thus, if Bertha were to remain behind of her own volition, it would violate her right if her companions were to bind her legs so that she could not flee. When a person is willing to fulfil a duty to incur a harm, it is only in the comparatively rare cases in which he cannot do it on his own that it is permissible for third parties to impose the harm on him. Except in such cases, only when a person resists her duty to incur a harm does it become permissible for third parties to enforce the duty by imposing the harm. But in such cases the justification for imposing the harm is no longer a matter of enforceable duty only. For the justification has now become in part a liability justification. In refusing to fulfil his duty to incur a harm, the person has forfeited his right against third party intervention. His duty is enforceable and, by refusing to fulfil it himself, he makes himself liable to proportionate harm when that is necessary to enforce his compliance. This can be seen in the fact that, if he refuses to incur the harm that it is his duty to incur, it can be permissible, in order to impose that harm on him, to harm him by even more than that amount, if necessary. The justification for the infliction of the necessary additional harm is a liability justification, which also now overlaps with the duty-based justification in overdetermining the justification for causing him the harm it is his duty to incur.

It is worth noting that some instances in which a person has an enforceable duty to bear a harm are also cases in which that person has a right not to suffer that harm. Consider, for example, Tadros's example in which 'I am the only person who can save the rest of the human race from being killed but I can only do so at the cost of my life'. Suppose the circumstances are that Martians will blow up the planet unless I let them kill me. I could evade them in my spaceship and join my friends on Jupiter, but I have an enforceable duty to remain and allow them to kill me. My having this duty is, however, compatible with the retention of my right against the Martians that they not kill me. (I do not, of course, have a right not to be restrained if I try to make a getaway in my spaceship. That I lack this right is what it means to say that my duty is enforceable.)

To me, the most interesting part of Tadros's commentary is his discussion of the case he calls 'Double Hit Man 2'. I have been puzzled by cases of this sort for some years.[9] Double Hit Man 2 has seemed to me to be a case in which Fred is liable, in the sense in which I use that term, to be killed as a means of saving Wayne from Evelyn's assassin, and Evelyn liable to be killed as a means of saving Wayne from the threat for which Fred is responsible. It has seemed to me, that is, that Fred and Evelyn have both forfeited their right not to be killed in defence of Wayne. I have found this intuition worrying, though, because one condition that I have thought necessary for liability to be killed is missing: namely, that the threat to be averted by killing a person should be one for which he is to some degree causally and morally responsible. I have, in other words, been concerned about being committed to accept the implications of the idea that a person can be liable to be killed as a means of averting a threat for which he is in no way responsible.

But Tadros may have come to my rescue. For he suggests that the way to understand this case is to see both Fred and Evelyn as having an enforceable duty to avert the threat for which he or she is responsible. Since neither can fulfil this duty but each can fulfil the other's, each has a duty to authorize the other vicariously to fulfil his or her original duty. Fred, for example, has a duty to permit Wayne to kill him to enable Wayne to defend himself from Evelyn's assassin, while Evelyn has a duty to permit Wayne to kill her to defend himself from Fred's assassin. These are enforceable duties. Hence Wayne is justified in enforcing these duties by using the killing of Fred and Evelyn as a means of saving himself from the assassins.

[9] For examples that raise some of the same issues but are, for various reasons, less effective than Double Hit Man 2, see Jeff McMahan, 'Self-Defense and Culpability', *Law and Philosophy* 24 (2005), pp. 751–74.

Tadros says that this is a case in which Fred and Evelyn have made themselves liable to be killed. But what he means, of course, is that each lacks a right not to be killed, not necessarily that each has forfeited his or her right not to be killed. I have been reluctant to accept that a person can *forfeit* his right not to be killed as a means of averting a threat for which he is in no way responsible. Yet I share Tadros's intuition that it is permissible to kill each of these people as a means of saving Wayne from the other's assassin. The explanation that Tadros offers of why it is permissible for Wayne to kill both Fred and Evelyn as a means of saving himself does not claim that either of them forfeits a right; rather, it claims that each has a duty to serve as a shield against the bullet fired by the other's assassin. If each simply acquires a duty rather than forfeiting a right, this explanation involves no appeal to the notion of liability as I understand it. Thus, if Tadros's explanation is correct, it solves my problem, for it shows why it is permissible for Wayne to kill Fred and Evelyn even though neither has forfeited the right not to be killed as a means of protecting Wayne from a threat posed by someone else.

There are, however, reasons for scepticism about this explanation. Suppose that Fred's assassin's gun jams and the assassin, in frustration, throws it out the window. Fred poses no threat to Wayne. He therefore has no duty to protect Wayne from a threat he poses, and hence no duty to make an agreement with Evelyn to fulfil her duty in exchange for her fulfilling his. If that is correct, he has no enforceable duty to submit to being killed as a means of saving Wayne from Evelyn's assassin. Yet my intuition is that it is still permissible for Wayne to use Fred as a shield against Evelyn's assassin, thereby killing him or making an essential contribution to his being killed. It was, after all, only a matter of luck that Fred's assassin's gun jammed. If it remains permissible for Wayne to use Fred as a shield, this supports the claim that I have in the past been tempted to make about cases of this sort – namely, that Fred makes himself liable to be killed (that is, forfeits his right not to be killed) by virtue of his responsibility for an *attempt* at murdering Wayne. But that presupposes that a person can be liable to be killed as a means of averting a threat for which he is neither causally nor morally responsible. And this, of course, raises the question of what the *exact* basis of Fred's liability is. Is it, as I just suggested, the culpable attempt? If so, is it essential that the attempt was very recent? Or might Fred be liable to be killed now as a means of saving Wayne on the basis of an attempt he made on Wayne's life a year ago? Might he be liable because he made a recent attempt on someone else's life? Or an attempt on someone else's life a year ago? It is in part to hold the line against the expansion of liability in these ways that I have sought to defend the view that for a person to be *liable*

to be harmed or killed as a means of averting a threat, he must be to some degree morally responsible for that particular threat.

It would be defensible, however, for Tadros simply to embrace the implications of his explanation. He can defensibly claim that it makes a significant moral difference whether Fred really does threaten Wayne's life. He can accept that, if the assassin's gun jams, Fred becomes, in the relevant sense, an innocent bystander, so that Wayne must allow himself to be killed by Evelyn's assassin rather than using Fred as a shield. But I remain sceptical of that conclusion.

Tadros argues further that if, as he suggests, people can become liable to be harmed, or acquire a duty to incur a harm, without being responsible for a wrong or a threat of wrongful harm, this can make it easier to justify harming or killing unjust civilians as a side effect of military action in a just war. He argues, more specifically, that civilians in a state fighting an unjust war can acquire a duty 'to bear some of the costs of the war even if they are not morally responsible for war', if they have causally contributed to their state's ability to fight or if they 'benefit from the state that poses the threat'. I think that there is an element of truth in this but that the full truth is subtler and more complicated.

Causal contribution and benefit are separate factors. I assume that Tadros means that either on its own could ground a duty to incur harm as a side effect and that both together could ground an even stronger duty. I will consider causal contribution first.

The example that Tadros gives of causal contribution without moral responsibility is participation by civilians in their state's economic activities – for example, by working at a job – through which they create the wealth necessary for their state to fight an unjust war. I think it is unreasonable to suppose that this could be a basis for a duty to suffer the side effects of military action. If doing x can be a basis of a duty to incur a harm if one's state fights an unjust war, it seems that there must be a moral reason, though not necessarily a decisive one, not to do x. But it is true of most people in the world who are employed that the work they do makes a tiny contribution to their state's ability to fight an unjust war. It is doubtful that that gives them *any* reason to abstain from working. One might claim that they have *a* reason but that it is easily outweighed by the reasons they have to work at a job. But if the reason is outweighed in the case of virtually every employed person in the world, the fact that these people have been employed and have thus contributed to the economy of their society cannot ground a significant duty to incur costs if their state fights an unjust war. Indeed, if these people are morally justified in working at a job, I think that in itself exempts them from a duty to bear side effect harms simply because they have worked. It is worth noting, moreover, that if people

can acquire a duty to incur costs simply by contributing to the economic strength of a state that fights an unjust war, it seems to follow, because of global economic interdependence, that when a state fights an unjust war, not just its own citizens but also a high proportion of the rest of the people in the world acquire a duty to incur harms as a side effect of military action against that state. I do not find that plausible.

Next consider Tadros's claims about benefit. Here I think we need a further distinction – namely, between benefiting from the functioning of a state and its institutions and benefiting from the state's unjust war. Tadros's claims are about the former. He writes, for example, that 'The fact that I contribute to my state and benefit from it in return can ground a duty to correct some of its actions even if I lack responsibility for the conflict. And this can affect the permissibility of harming me' (p. 276). The idea here is that the duty to correct the state's wrongful acts can be fulfilled, indirectly, by bearing certain harms caused as a side effect of military action against the state.

I do not accept that merely benefiting from the functioning of a state, even if one has made a causal contribution to its ability to function, can be the basis of a duty to incur harms as a side effect of just military action against that state. The idea that it can implies, among other things, that if the state fighting the unjust war is one of the major economic powers, a high proportion of the people in the world have such a duty, since these people benefit from many of that state's economic activities and have contributed to its wealth in various ways, through trade, labour, lending, the provision of services, and so on.

It is more plausible to suppose that a person may acquire duties by benefiting from an unjust war. What seems most plausible to me is that, if a person has acquired wrongful gains as a result of an unjust war, he has a duty to surrender those gains if his doing so would benefit the victims of the war – that is, those whose losses have made his gains possible. He might have a duty to surrender his gains by paying reparations to the war's victims, or a duty to bear losses caused as a side effect of efforts to prevent further losses to the victims – for example, deprivations caused by economic sanctions. The important point is that the costs that it may be a civilian's duty to incur solely as a result of benefiting from an unjust war cannot exceed what is necessary to cancel his or her wrongful gains. In the great majority of cases, those costs fall well short of the harms that civilians typically suffer as a side effect of military operations.

It therefore seems to me that the fact that the designer and builder of Tadros's remote control plane has benefited from his work is irrelevant to the permissibility of the bystander's diverting the plane so that it will kill the designer rather than himself. This is, in my view, a straightforward case of liability derived from responsibility for a threat

of wrongful harm. This person has built an enormous object using a new, untested technology and sent it hurtling through the air above an area where there are people on whom it could crash. If it is now unavoidable that it will kill someone, it is permissible as a matter of justice in the *ex ante* distribution of harm to ensure that it will be the person responsible for the threat who is killed rather than an innocent bystander, assuming that other relevant considerations are equal. This would be true even if the designer had not benefited from his work on the plane at all – even, indeed, if he had incurred losses in building it.

This case does not, however, seem to have any obvious implications for civilian liability. The ways in which civilians ordinarily contribute to the threats posed by an unjust war are much less significant than the contribution that the designer and builder of a defective plane make to the threat it poses.

Suppose one still finds some intuitive plausibility in Tadros's contention that civilians who have causally contributed to a state's ability to fight an unjust war and have benefited from their relations with that state thereby acquire a duty of some sort with respect to that state's unjust war. What seems to me most plausible is that their duty would be to oppose the unjust war, particularly if they are citizens of the state. Suppose, at least for the sake of argument, that that is so. As with other duties to incur costs or harm of the sort that Tadros discusses, this is a duty that they can fulfil on their own. It is only if they fail to fulfil this duty that they might become liable, as individuals, to suffer certain side effect harms, though only harms that would be proportionate in relation to the seriousness of the duty they have neglected. If, by contrast, they take reasonable action to oppose the unjust war, they thereby fulfil their enforceable duty and exempt themselves from liability to bear harms that are side effects of military action in a just war against that state.

mcmahan@philosophy.rutgers.edu