

Liability and Necessity

Helen Frowe, Kent and ELAC

Introduction

The notion of *liability to defensive harm* has become prominent in the self-defence and just war literature.¹ Roughly, to say that a person is liable to a harm is to say that inflicting that harm upon them would not wrong them: they are ‘morally open’ to the harm. Harming them does not violate or infringe their rights because they lack a right against having the harm inflicted upon them.

But a division has emerged between those who hold what I will call *internalist* accounts of liability, and those who hold what I will call *externalist* accounts.² Internalists think that liability to defensive harm is contingent upon necessity; externalists think that one can be liable even to unnecessary defensive harms. Internalism faces the problems of granting even culpable aggressors rights to compensation and rights of counter-defence in these cases. Externalism faces the problem of explaining the moral worseness of killing an aggressor who is less than fully culpable, compared to killing a fully culpable aggressor. It also faces the problem of reconciling the overall impermissibility of unnecessarily harming an aggressor with denying that aggressor a right to fight back.

In this paper, I argue that internalism is false: it’s not the case that one can be liable only to necessary defensive harms. But it *is* the case that only instrumental harms can count as defensive. As such, only these harms can be captured by an account of liability to

¹ See, for example, Jeff McMahan, *Killing in War* (Oxford: OUP, 2009), Cecile Fabre, ‘Guns, Food and Liability to Attack in War’, *Ethics*, Jon Quong, ‘Liability to Defensive Harm’, *Ethics*, Helen Frowe, ‘A Practical Account of Self-Defence’, *Law and Philosophy*, (2009).

² Quong and Firth use the labels ‘instrumentalist’ and ‘non-instrumentalist’ accounts to refer to internalism and externalism respectively. But I’m going to stick with my labels, because they better capture the necessity element. After all, in a case where Victim can escape a threat Attacker poses by killing Attacker, or by running away, killing Attacker is still instrumental in averting the threat Attacker poses. But it’s cases like this in which people like McMahan want to say that Attacker isn’t liable to be killed because killing him violates the necessity condition - it’s not the *least harmful means* of averting the threat Attacker poses. As will become apparent, I think that instrumentality is key in this debate.

First draft: please do not quote or circulate.

defensive harm. I argue that by recognising this, along with a distinction between defensive action and defensive harm, we can see that neither internalist nor externalist accounts need be troubled by cases of insufficient defence, as some writers have supposed. If an aggressor is indeed liable to suffer harm in insufficiency cases, as I think he is, we have to explain this liability on grounds other than the aggressor's liability to defensive harm. But I will argue that this is true of both internalist and externalist accounts, and so this result doesn't give us a reason to prefer one account over the other.

In light of this argument, we can see that a recent attempt by Jon Quong and Jo Firth to undermine internalism by looking at insufficiency cases and impossibility cases – where the victim has no opportunity or ability to fight back – is misguided. I also argue that Quong and Firth's moral-luck based objection to internalism fails, and that their own pluralist view produces largely unconvincing results.

In its place, I argue for what I call an instrumentalist account of liability to defensive harm. Like externalism, it that holds a person liable to defensive harm on the grounds of their moral responsibility for an unjust threat. However, I take seriously that this is liability *to defensive harm*, and that this must mean only instrumental harm: that is, harm that is capable of defending a person. I argue that in order to distinguish between fully culpable and less than culpable aggressors, the necessity constraint that helps determine the overall permissibility of using force must be sensitive to the degree of an aggressor's moral responsibility. Whether a use of force counts as the least harmful means must involve a judgement about how harms to the affected parties are weighted. This enables my view to capture the intuition that Victim ought to bear more cost to avoid harming a minimally responsible aggressor than a culpable aggressor.

I. Internalism and Externalism

First draft: please do not quote or circulate.

Internalism holds that one can be liable to a defensive harm only if the infliction of that harm is necessary for averting a threat – that is, if it is the least harmful means of averting the threat. We can think of this as the view that liability has an internal necessity condition. Jeff McMahan is an internalist, holding that “the assignment of liability is governed by a requirement of necessity. If harming a person is unnecessary for the achievement of a relevant type of goal, that person cannot be liable to be harmed.”³

Internalists distinguish liability from desert by noting that a hallmark of desert is that when a person deserves a harm, that is itself a reason to give it to them. As McMahan puts it, it’s impartially good that people get what they deserve: if a person deserves a harm, inflicting the harm “is an end in itself.”⁴ The same is not true of harms to which people are merely liable. Such harms are bad “not only for those who suffer them but also from an impersonal point of view.”⁵ Desert entails liability, but liability does not entail desert.

As externalism is normally understood, it holds that a person’s liability to defensive harm is independent of whether or not inflicting that harm upon them will avert a threat. Generally, externalists think that liability is ‘backwards-looking’, fixed by facts about, for example, the person’s moral responsibility for posing or contributing to an unjust threat. (In this paper, I’ll treat this as the basis of liability in the externalist account.) It is thus immune to thoughts about the necessity of harming a person, since this does not influence whether the person is morally responsible for an unjust threat.

Externalists can also distinguish between liability and desert. To say that a person deserves a harm implies a positive reason to inflict it upon her. As I understand liability, to say that a person is liable to a harm means only that a usual reason not to harm them – that they have a right not to be harmed – is absent. But to say that a person is liable to harm

³ McMahan, *Killing in War*, p. 9

⁴ KIW p. 8

⁵ KIW, p. 8

First draft: please do not quote or circulate.

entails no claim that it is desirable to inflict harm on her: it is not a claim that it is good that she be harmed, in the way that it is generally good that people get what they deserve.

II. The importance of liability

It's likely that in many cases of self-defence, the internalist and externalist will give the same headline judgement concerning whether Victim is permitted to defend himself. Consider

Lucky Escape:

Lucky Escape: Murderer is shooting at Victim to try to kill him because he dislikes Victim. He chases Victim to the edge of a cliff. Unbeknown to Murderer, Victim has both a gun and a parachute. He can thus save his own life by either (a) jumping to safety, using no force against Murderer, or (b) shooting and killing Murderer.

In *Lucky Escape*, force is unnecessary for averting a threat because it is not the *least harmful means* the victim has for averting the threat. Therefore, internalists and externalists will both say that it is impermissible for Victim to kill Murderer. But they will differ on whether Murderer is liable to be killed. The internalist will argue that Murderer is not liable to defensive harm. But the externalist will disagree: Murderer is still morally responsible for an unjust threat, and it's this that makes him liable to defensive harm on the externalist account. Killing him would be wrong, because it fails the necessity condition. But necessity is an external constraint on the *overall permissibility* of inflicting harm, not constitutive of liability itself.

First draft: please do not quote or circulate.

Why does this difference in liability matter, if not for the purposes of determining what Victim may do in self-defence? Well, liability has implications for at least two important issues. Usually, if one inflicts a harm upon a non-liable person, that person is entitled to compensation for the harm she suffers. So, if Murderer is not liable to defensive harm, but Victim nonetheless inflicts harm upon him, Murderer seems to have a claim that Victim compensate him.

More importantly, one is typically permitted to defend oneself against the infliction of a harm to which one is not liable. So, Murderer's liability may be crucial to the question of what he may do to Victim should Victim use defensive force against him. It's possible, on the internalist view, that by wrongly trying to kill Victim in circumstances where Victim is able to escape, Murderer acquires a moral permission to kill Victim if Victim uses even proportionate force against him. This, in turn, could ground a permission or even an obligation for third parties to assist Murderer. This is pretty counter-intuitive.

There's a third implication of the internalism / externalism debate that we should note [although I won't have time to say much about it here]. The role we grant to necessity can determine whether we adopt either a narrow or a broad view of liability. A person is *narrowly* liable to be harmed if she is liable to be harmed only to avert the particular unjust threat for which she is responsible. A person is *broadly* liable to be harmed if, once she is morally responsible for posing an unjust threat, she is liable to harm to avert any unjust threat, provided the harm we inflict upon her is proportionate to the threat for which she is responsible.

This distinction has significant implications for killing in war. We can illustrate it using the following case:

First draft: please do not quote or circulate.

Alley: First Shooter is shooting through a basement window at innocent Victim, maliciously trying to kill him. In an independent attack, Second Shooter is shooting at Victim from the roof. Victim can hide from Second Shooter, but not from First Shooter. However, he can shoot Second Shooter, and Second Shooter's body will then fall from the roof and block First Shooter's line of fire, saving Victim's life.⁶

The narrow view of liability holds that Second Shooter is liable only to harm that averts the threat for which he is responsible. Since he's not responsible for First Shooter's attack, he is not liable to be killed to avert it (even if killing him is necessary to avert it). But the broad view of liability holds that Second Shooter can be liable to be killed so that he blocks First Attacker's line of fire. Once he is liable to be defensively killed by Victim, Victim can kill him in defence against another threat to his life.⁷

Internalism is neutral between narrow and broad liability. Internalists need only require that a harm be instrumental in averting *a* threat – they needn't say that it must avert the very same threat for which the target is responsible.⁸

Externalism is compatible with broad liability. Since Second Shooter can, on the externalist view, be liable to be killed even if killing her serves *no* purpose, she can surely be liable to be killed for *any* purpose – as on the broad view. But externalism is not compatible

⁶ I've adapted this from a set of cases in McMahan, 'Self-Defense and Culpability', *Law and Philosophy*, 24, (2005): 751 – 774, p. 757

⁷ These two kinds of liability correspond to views expressed by McMahan and Hurka in their debate about proportionality in war. Hurka thinks that once a country makes itself liable to attack, it also makes itself liable to force aimed at achieving additional goods, such as disarmament, even if it would otherwise be disproportionate to pursue disarmament by means of war (broad view). McMahan, in line with the narrow view, argues that only those ends that do themselves warrant pursuit by military force (i.e. serious rights violations) can be factored into the proportionality calculation. The country is liable only to force necessary for the correction of those particular wrongs that made war permissible. See McMahan, 'Just Cause for War', *Ethics and International Affairs*, 19, 3, (2005): 1 – 21, and Hurka, 'Proportionality and the Morality of War', *Philosophy and Public Affairs*, 33, 1 (2005): 34 - 66

⁸ Some internalists do say this – McMahan ties liability to averting the threat for which the attacker is responsible. But they don't *have* to say this.

First draft: please do not quote or circulate.

with narrow liability. If one doesn't tie liability to averting a threat at all – as on the externalist account – then one cannot tie it to averting a particular threat, which is what the narrow view of liability requires. If, in order to be liable to be killed, the target of force must be morally responsible for the particular threat one is trying to avert, this will entail that she can be liable to be killed only if killing her aims at trying to avert a threat. The narrow view of liability doesn't require *necessity* for liability – that force aims at averting a threat doesn't mean it's the least harmful means of averting it. But it does require *instrumentality* – that the force be *a* means of averting a threat – which seems to make it incompatible with externalism's exclusively backwards-looking focus on responsibility for an unjust threat.

I won't attempt to settle this question here: McMahan's paper shows that this is a complicated problem and that neither view is obviously correct, although he ultimately opts for the narrow account. But it's worth noting that which way one goes with respect to internalism and externalism could have a bearing on this debate too.

III. Insufficiency cases

Lucky Escape illustrates one type of necessity problem: it is a case in which force is not the least harmful means of avoiding or averting a threat. A different type of necessity problem involves the use of force that is unnecessary because it is *insufficient* to avert the threat the victim faces, and yet it is the maximum amount of force the victim has at her disposal. Quong and Firth explore the problem of insufficient force with the following case, *Rape*:

Rape: Eric is in the midst of culpably rape Fran. Eric is much bigger and stronger than Fran, and consequently there is nothing she can do to stop him from continuing to rape her. While being raped, Fran tries to resist and in doing so threatens to break Eric's wrist, thought this will do nothing to stop the rape

First draft: please do not quote or circulate.

from occurring. The only way Eric can stop Fran breaking his wrist is to quickly break her wrist first.⁹

Suzanne Uniacke has pointed out that a harm must have some chance of averting a threat to count as *necessary* defence.¹⁰ If so, it looks like the moderate harm Fran can inflict fails the necessity condition.

Rape raises not only the question of whether Fran may break Eric's wrist when doing so won't avert the rape, but also the question of whether Eric may do anything to stop her. Quong and Firth say that the internalist account will tell us that because breaking Eric's wrist "stands no chance" of averting the rape, Eric cannot be liable to it. Thus, internalism must permit Eric to defend himself against the infliction of the broken wrist by breaking Fran's wrist first.

As Quong and Firth say, that seems a pretty dreadful result. A natural response that Quong and Firth consider is to say that Eric is liable to punitive harm, and that Fran is thus permitted to break his wrist on these grounds. But, Quong and Firth say that this isn't good enough – the correct account of liability to defensive harm should itself contain the resources to avoid permitting Eric to break Fran's wrist. It shouldn't need to be supplemented by thoughts about punishment or other justifications for harming.

(i) Defensive action and defensive harm

I think that to explain what's going on in this sort of case, we need to recognise a distinction between defensive action, and defensive harm. Whether an action is defensive is determined by the agent's beliefs and intentions. Consider *Sprinkler*:

⁹ P. 689

¹⁰ Suzanne Uniacke, 'Self-Defence, Just War and a Reasonable Prospect of Success', Helen Frowe and Gerald Lang (eds.), *How We Fight: Ethics in War*, (Oxford: OUP, 2014)

First draft: please do not quote or circulate.

Sprinkler: Gunman is just about to shoot Victim through Victim's kitchen window. Victim has no idea that Gunman is there, but since he's off to bed, he turns on his powerful garden sprinkler. A jet of water hits Gunman in the eye as he pulls the trigger, causing his bullet to miss Victim.

I don't think we should say that in turning on his sprinkler, Victim acted defensively, because he lacks the requisite beliefs about there being a threat and lacks any defensive intent. But I do think the harm he inflicts on Gunman is a defensive harm, because it is in fact a harm that defends his life.

Now imagine that Victim knows that Gunman is about to shoot him, but he has no sprinkler and there's no time to do anything else to stop Gunman from killing him. Victim spits at Gunman. He has no expectation that this will avert the threat – he correctly believes it will do nothing at all to deter Gunman or deflect the bullet. Is Victim acting in self-defence when he spits at Gunman? I don't think he is, because not just anything that one does whilst knowing that one's life is in danger can count as acting in self-defence. The category of defensive actions must be demarcated in some way, and the most plausible way to do this is to look at the agent's beliefs and intentions. And since spitting at Gunman will do nothing to avert the threat, it doesn't count as a defensive harm either. I mentioned earlier Uniacke's claim that a harm must have some chance of averting a threat if it is to count as necessary. But we can make a stronger claim: a harm must have some chance of averting a threat if it is to count as *defensive*. Spitting at the gunman in the *Sprinkler* case is not an unnecessary defensive harm – rather, it is not a defensive harm at all.

First draft: please do not quote or circulate.

It might be helpful to think about a roulette case to see how defensive action and defensive harm can come apart. If all the chambers of your gun are empty when you point it at my head, killing you cannot avert a threat to me (because there is no threat) and thus to kill you is not to inflict a defensive harm. But I can still be sensibly said to be acting in self-defence when I kill you if I believe that killing you will avert a threat to my life.¹¹ And, I think that you're still *liable* to be killed by me in this case because it's your fault that I think I have to kill you to avert a threat to myself. You're just not liable on the grounds of defence. You can't be, on either an externalist or internalist view. Externalism requires moral responsibility for an unjust threat, and here there is no threat. Internalism requires both the unjust threat and that harming you be necessary for averting that threat. Again, since there's no threat, there are no internalist grounds of liability to defensive harm. But it's not only defensive harm to which a person can render herself liable to through her responsible behaviour. She can render herself liable to other types of harm as well.

I suggest, then, that for an agent to count as acting self-defence, she must believe that her action has some chance of averting a threat. And for a harm to count as defensive, it must be capable of averting a threat.

What does this tell us about a case in which Fran correctly believes that breaking Eric's wrist will not avert the rape? In this case, Fran is not acting in self-defence, and breaking Eric's wrist is not a defensive harm. If Fran is nonetheless permitted to break Eric's wrist, this permission must again be captured by something other than an account of permissible defence, such as Eric's liability to punitive harm, or Statman-esque symbolic harm or harms that re-assert dignity.

¹¹ The cases aren't exactly analogous: in *Rape*, Fran is obviously correct in her belief that there is a threat – her mistaken belief is her belief that a broken wrist will be sufficient to avert that threat. In the *Roulette* case, the mistaken belief is that there's a threat that killing you can avert. But I think the comparison still supports the general idea that I can plausibly be said to be acting defensively even in cases where my actions cannot avert a threat – indeed, even when there is no threat, as in the *Roulette* case – and that an aggressor can be liable to harm in these cases, even if they cannot be defensive harms.

First draft: please do not quote or circulate.

So, Eric might not be liable to having his wrist broken *as a defensive harm* when breaking the wrist has no chance of averting the rape (although of course, he remains liable on grounds of defence to harm that *would* avert the rape). But he can be liable to a broken wrist on other grounds – for example, that it’s Eric’s fault that Fran believes that she must break his wrist to avert a serious harm to herself. If so, both an internalist and an externalist can explain why Eric may not defend himself against the broken wrist – they just can’t do so by invoking Eric’s liability to defensive harm.¹²

But contrary to what Quong and Firth claim, this result is unavoidable. It’s a mistake to think that “the correct account of liability to defensive harm should not require appealing to external moral considerations” to explain why Eric is liable to have his wrist broken in these circumstances.¹³ Rather, *only* something other than an account of liability to defensive harm could possibly explain this, because the broken wrist is not a defensive harm. It can’t be, because it can’t defend Fran. A harm need not be the least harmful means available for averting a threat to count as defensive, but it must be *a* means of averting threat to count as defensive.

Quong and Firth contend that it must be an account of defence that captures Fran’s permission to break Eric’s wrist, because “there is something deeply troubling about the proposal that an innocent person under threat has no liberty whatsoever to *defend* herself... there is something problematic about the view that morality can render an innocent person

¹² Obviously, referring to things like the prospect of averting the threat, and whether force has any chance of averting the threat, is to invoke probabilities. It might be tempting to say that if a harm did not in fact avert a threat, then it never had any prospect of averting the threat, and thus all unnecessary harms fail to be defensive. I don’t think that’s right – we can perhaps extrapolate from previous experience to give a more objective sense to probabilities – if breaking an attacker’s wrist has sometimes averted rapes, we might think that’s enough to make it a harm that has some prospect of averting rapes. If scratching an assailant has never averted a rape, we might say that scratching has no prospect of averting rapes. I don’t have space here to get into how we decide whether a given force has any prospect of averting a threat. But it won’t matter too much for determining Eric’s liability (although it may determine whether he is liable on grounds of defence or some other ground). As long as Eric is morally responsible for making Fran believe that harming him is necessary to avert a threat to herself, he’s liable to bear the (proportionate) harm she inflicts as a result of that belief.

¹³ Quong and Firth, p. 690

First draft: please do not quote or circulate.

defenceless. (emphasis in original)¹⁴ But whilst it might be unsettling, it's simply true that morality can prohibit an innocent person's defending herself. If Betty can defend herself against Albert only by detonating a bomb that kills all six of her neighbours, morality renders Betty defenceless in this case. The only way she can save her life is by killing Albert, and morality does not allow this when she will also kill six other people.¹⁵ It isn't the case that the correct account of self-defence must always allow innocent people to defend themselves.

IV. Impossibility cases

Quong and Firth also seek to undermine internalism by arguing that it generates implausible results in cases like *Unjust Aggression*:

Unjust Aggression: Carla is preparing to unjustly attack Dan in order to steal some of his property. She draws up two possible plans of attack. Under the first plan, she will wait until the next time he invites her over, slip a sedative in his coffee, and steal the goods. Under this plan, Dan will be left relatively unharmed (except for the stolen property). But there is a reasonable chance that he will be able to successfully thwart her since she can't get hold of a very strong sedative, and he might notice her slipping the sedative into the coffee. Under the second plan, Carla will break into Dan's house while he is watching TV, kill him using

¹⁴ Quong and Firth, p. 690

¹⁵ We might, following Suzanne Uniacke, distinguish between cases in which Betty has a right to defence that she cannot exercise, and those where she lacks the right of defence, but (a) the upshot of both characterisations of rights is that morality prohibits innocent Betty's defence against a culpable threat, which is what Quong and Firth object to, and (b) on Uniacke's view, neither Fran nor Betty need to be said to lack a right of self-defence. Fran has a right to kill Eric in self-defence – she just can't exercise it because she lacks the strength, and Betty has a right to kill Albert in self-defence – she just can't exercise it because of the external constraints imposed by the presence of bystanders.

First draft: please do not quote or circulate.

a gun and steal the goods. Under this second plan, Carla is certain to succeed and kill Dan.¹⁶

Quong and Firth claim that because in the second scenario, “any attempts at self-defence by Dan are certain to be instrumentally ineffective”, the internalist account must pronounce that Carla is not liable to defensive harm. Internalism thus “yields the very odd result that Carla can ensure moral immunity from defensive harm by choosing to act in a far more harmful and unjust manner.”¹⁷ I’ll call this an *impossibility case*, since the idea behind the objection is that by choosing means that make it impossible for Dan to defend himself, the internalist must say that Carla is not liable to defensive harm.

What Quong and Firth say about this case (and some of the other cases they discuss) suggests to me that they just haven’t understood internalism, or at least that they’re not addressing the most plausible version of it. The internalist view will not hold that Carla is immune to defensive harm because she attacks Dan with overwhelming force that he cannot defeat. Rather, it will hold that that Carla is liable to be killed by Dan because (let’s assume) killing her would avert the threat, and nothing less than killing her will avert the threat. If Dan could kill her, then, this would be the least harmful means of averting the threat to his life. The fact that he can’t kill her doesn’t show that killing Carla isn’t necessary, and doesn’t show that she isn’t liable to be killed on the internalist view. Whether or not a given amount of force will stop Carla from killing Dan is not determined by Dan’s capabilities.

There’s a difference between whether proportionate defence is *possible*, and whether proportionate defence would *work*, where ‘working’ means averting the threat. Internalists reject liability to defensive harm in cases where (a) one can escape the threat using less harmful means, as in *Lucky Escape*, or (b) even if one used proportionate defensive force,

¹⁶ Quong and Firth, p. 688

¹⁷ P. 688

First draft: please do not quote or circulate.

that force would not work – it would not avert the threat. This is why McMahan says that non-combatants are generally not liable to be killed – not because killing them is impossible, but because *even if one killed them*, this would not avert the threats for which the non-combatants are responsible. As he puts it, “a person is liable to be harmed only if harming him will serve some further purpose – for example, if it will prevent him from unjustly harming someone else.”¹⁸

To see the difference between this sort of case and *Unjust Aggression*, consider *Unjust Poison*:

Unjust Poison: Carla has been slowly poisoning Dan over several weeks. Dan has already consumed a lethal dose of the poison and will certainly die no matter what.

In this case, killing Carla will serve no purpose and in *this* case the internalist will say that Carla is not liable to defensive killing. But this is because killing Carla would not *work*, irrespective of whether killing her is possible. In these *ineffective* cases, in which there is no amount of force that could avert the threat for which she is responsible, no harm inflicted on Carla can be instrumental. Thus, no harm inflicted upon Carla can be defensive – and that holds for both externalist and internalist accounts. But internalists needn't reject liability in cases where killing the aggressor would be proportionate and would avert the threat, but defence is not possible because the victim lacks the strength or means to inflict the necessary harm.

¹⁸ Another way in which we can see that Carla is liable to be killed is by thinking about what a third party could do. It's obvious that a third party could shoot Carla to save Dan, and most people think that a third party's right of other-defence to kill Carla must derive from Dan's right of self-defence against Carla. If Carla really did have moral immunity against being harmed, it's hard to see why a third party would be permitted to kill her.

First draft: please do not quote or circulate.

Of course, there's a different sense in which the necessity condition *is* tied to the defender's capabilities, illustrated by *Murder 3*:

Murder 3: Albert is trying to kill Betty. He'd be stopped by punch on the nose.
 But Betty is too short to punch Albert on the nose. The only way Betty
 can stop him is by stabbing him in the chest.

We should not understand the necessity condition as it applies to individuals engaged in defensive actions as requiring that one use the least amount of force that would in fact avert the threat. The necessity condition requires that individuals use the least harmful means available to them (and is of course governed by proportionality).¹⁹ But as long as there is an amount of harm that would prevent the attack – irrespective of whether the victim or anyone else can inflict that harm – then the least amount of force that would avert the threat is necessary for averting it. If it is also proportionate, then an internalist can hold that the target is liable to that force – again, irrespective of whether the victim or anyone else can inflict it.

The only cases that really distinguish internalism from externalism, then, are the *Lucky Escape*-type cases, in which harm against the aggressor is instrumental, but is not the least harmful means of averting a threat.

V. Moral luck and liability

I used to think that an important strike against internalism is the way in which it makes Murderer's liability dependent on a range of variables.²⁰ For example, in the event that it would be risky for Victim to use his parachute (perhaps the parachute is old and unreliable, or the cliff very high), it seems to me permissible that Victim kill Murderer rather than risk his

¹⁹ I come back to this in Section VII below.

²⁰ See Frowe, 'Non-Combatant Liability in War', Frowe and Lang, *How We Fight: Ethics in War* (Oxford: OUP, 2014)

First draft: please do not quote or circulate.

own life by jumping. And, I don't think the risk to Victim would have to be very high to make this permissible: I think that if Victim is just not sure whether he will be able to jump to safety because the wind is quite strong, then he is not required to jump. He may instead shoot Murderer to save his own life.

But we might think that there is something troubling with imagining Murderer's liability to defensive harm to pop in and out of existence on the basis of Victim's confidence in his parachuting skills, or the condition of his parachute. The internalist view generates the peculiar result that if he inadvertently attacks a world-class parachuter, who is utterly confident in his ability to jump to safety, Murderer is not liable to be killed. But, if he attacks a novice parachuter, or one who hasn't looked after his equipment – well, then he might be so liable. And yet in both cases Murderer is responsibly trying to inflict lethal harm upon an innocent person. It seems more plausible to think that Murderer's liability remains constant across these variations, whilst the overall permissibility of harming him changes. This captures the intuition that Murderer ought to have control over his liability to being harmed, but that there's less reason to think he ought to have control over whether harming him is all-things-considered permissible.

Quong and Firth argue that the role internalism allows for moral luck is indeed a central reason to reject the internalist view, for precisely the kind of reason I have given – that the right not to be non-consensually killed is an important agency right, and its forfeiture “should not be influenced by the sort of luck to which the [internalist] account is sensitive.”²¹ But I'm no longer sure that this kind of moral luck-based objection is very successful against the internalist view – rather, it might disappear if we think more carefully about the baseline against which this moral luck can operate.

²¹ Quong and Firth, p. 680

First draft: please do not quote or circulate.

Murderer is culpably and unjustly threatening Victim's life. The normative baseline here is that Victim may kill Murderer if he needs to. The sensitivity to moral luck, then, can serve only to *improve* things for Murderer: he gets *good* moral luck if it turns out that Victim doesn't need to kill him, because he then retains his immunity to being harmed. Given this, Murderer is very unlikely to make the kind of complaint that Quong and Firth make on his behalf – that his rights shouldn't depend on this kind of variable. If these agency rights protect our normative control over our lives, as Quong and Firth claim, Murderer is surely going to prefer a situation in which he retains his right against being non-consensually killed to one in which this right is forfeit. And there doesn't seem to be any other moral objection to allowing luck to be relevant to a person's liability when we do so only when it makes the agent whose rights are in jeopardy better off, and makes nobody else worse off. Doing so is certainly not objectionable from Murderer's perspective, which is the perspective from which Quong and Firth ground their objection.

Quong and First might reply that allowing luck to play this role does make someone else worse. It makes Victim worse off. He is not worse off because he must now suffer some harm – he can still avoid that by using the parachute. But he's *morally* worse off because if he uses force against Murderer, Murderer is permitted to fight back or to claim compensation. But this reply threatens to beg the question against the internalist. Whether or not this forms the basis of a legitimate complaint from Victim is precisely what is at issue between the internalist and the externalist. The internalist thinks Victim *should* be morally worse off if he uses unnecessary force against Murderer, because he wrongs Murderer in that case.

If we want some further argument for why we should adopt the baseline I have suggested, against which luck serves only to improve Murderer's moral standing, we can point out that the situation in which Victim can save his life only by killing Murderer is the very situation that Murderer is trying to bring about. Presumably Murderer doesn't *want*

First draft: please do not quote or circulate.

Victim to have any other escape route - he isn't trying to kill him whilst secretly hoping that Victim can get away. The fact that he's trying to make it the case that Victim cannot get away means that we don't treat Murderer unfairly if we take that scenario - in which Victim cannot escape and thus needs to kill Murderer - as the normative baseline. But nor do we treat Murderer unfairly if, as it turns out, moral luck means that killing him is unnecessary and he gets to retain his right not to be killed.

VI. The pluralist account

Quong and Firth argue that we should reject internalism in favour of their pluralist account of liability, which combines externalism with a humanitarian duty to aid when one can do so at a low cost to oneself. According to this view, even when a person forfeits her right not to be killed by being morally responsible for an unjust threat, she retains a general humanitarian right to be aided when aiding her will impose only a low cost on others. From this right to be aided when the cost of aiding is low, we can derive a right not to be harmed when the cost of not harming is low.

So, a person who lacks a right not to be killed is only partly liable to be killed as a result. To be fully liable, such that killing them in no way wrongs them, they must also lack a right to this kind of basic humanitarian consideration. In cases where Victim can refrain from killing Murderer at only a low cost to himself, he violates Murderer's humanitarian rights if he nonetheless kills him. This captures Quong and Firth's view that Victim wrongs Murderer by killing him unnecessarily. But they argue that it also limits what Murderer can do to Victim if Victim uses force against him: "by breaching a low-cost duty such as the humanitarian duty, one becomes liable to only a low degree of force because this duty is only a duty to bear a low degree of cost."²² So, if Victim tries to kill Murderer, Murderer can use

²² P. 699

First draft: please do not quote or circulate.

some force to try to stop him, but only a pretty moderate amount, because any more would exceed Victim's liability. (Presumably any claim to compensation would be similarly limited, although Quong and Firth don't mention this.)

This is Quong and Firth's attempt to deal with the problem I mentioned in the introduction: reconciling the impermissibility of killing Murderer with denying Murderer a right to counter-defence. Externalism is, to a large extent, motivated by the intuition that Murderer shouldn't have a right of counter-defence. And yet, if also we want to say that Victim acts impermissibly in needlessly killing Murderer, we seem committed to the view that Victim fails in some sort of duty if he harms Murderer. If that's right, Victim must render himself liable to *something*, it seems, in light of this failure. Quong and Firth's compromise is to allow Murderer to use some force, but to ground this in a right to make Victim comply with his humanitarian duty rather than Murderer's right not to be killed, and limit the force Murderer can use to what would be proportionate to make Victim comply with this humanitarian duty.

Quong and Firth argue that although the pluralist account does not produce perfectly intuitive results in the cases they discuss, it does a significantly better job than the internalist account. But, as I have argued, they've largely misunderstood what internalists would say about the cases they discuss. Properly understood, internalism fares just as well or better than the pluralist view in many of their cases. For example, of *Unjust Aggression*, Quong and Firth say that "[o]n our account [Carla] cannot make herself immune to defensive harm by choosing to unjustly attack others in a manner that is certain to succeed. On the pluralist account, whether an attacker is fully liable to defensive harm in cases like *Unjust Aggression* depends on whether the attacker retains her humanitarian right, and this is determined by the cost to the defensive agent of refraining from harm, and not by the probability of the

First draft: please do not quote or circulate.

defensive harm being successful.”²³ They claim a difference of “real moral significance” compared to the internalist, because even though Dan is effectively defenceless, “it remains true that Carla is, as a matter of morality, liable to his defensive actions.”²⁴ But as I have argued, whether the internalist pronounces Carla liable to be killed does not depend on whether Dan can kill her. It depends on whether killing her would stop her from killing Dan. Assuming it would, internalism pronounces Carla straightforwardly liable to be killed. There’s no victory for the pluralist account here.

Quong and Firth argue that the pluralist view is also to be preferred to internalism in *Rape* because Eric’s humanitarian rights grant him protection only against serious harms. If a broken wrist is not a serious harm, Eric can’t use defensive force against Fran when she tries to break his wrist. But this is not really much of an improvement upon what they take internalism to say about *Rape*. All we need to do is increase the harm to a serious harm – say, a broken leg – and the problem re-appears. It seems to me no less objectionable that if Fran goes to break Eric’s leg, Eric gets to break Fran’s wrist to stop her and claim a right of self-defence against the woman he’s currently engaged in raping. But if Eric can defend himself against serious harms, that looks like what the pluralist view will allow him to do. Unlike Quong and Firth, I don’t find this “an intuitively satisfying” result.²⁵

Moreover, it’s not clear that Quong and Firth can restrict the amount of harm Eric can inflict to only mild harms. Consider *River*:

River: Swimmer is drowning in a shallow part of the river. At time *t*, Jogger can pull her out at the risk of getting sprained wrist. But Jogger doesn’t want to

²³ Quong and Firth, p. 699

²⁴ Quong and Firth, p. 700

²⁵ Quong and Firth, p. 699. Quong and Firth might try to argue that a broken wrist is still too serious a harm for Eric to be allowed to inflict it upon Fran to get her to comply with her humanitarian duty (although they don’t seem to want to call it a serious harm when Fran is inflicting it upon Eric). But we can decrease the harm to e.g. broken fingers – it still seems to me a poor result if the pluralist account grants Eric a moral permission to break Fran’s fingers as an act of self-defence while he’s in the middle of raping her.

First draft: please do not quote or circulate.

interrupt his run, so keeps going. Swimmer is swept downstream to where, at t_1 , Jogger can still pull her out, but now he'll risk a broken leg.

It seems to me that even if Jogger is not required to pull Swimmer out at the cost of broken leg at t , he is so required at t_1 . Jogger had an opportunity to fulfil his obligation to save Swimmer at low cost, but he wrongly failed to take that opportunity. This means that he can now be required to bear much greater costs to fulfil that original obligation. It also means that a third party could inflict those greater costs on Jogger to make him save Swimmer.

If this is right, it looks like Eric may actually use quite significant force against Fran to stop her from hurting him. Fran had an opportunity to fulfil her humanitarian obligation to refrain from harming Eric, but (according to Quong and Firth), she wrongly fails to take this opportunity when she tries to break his wrist. Even if initially, at t , she needed to bear only a low cost to fulfil her humanitarian duty, this needn't be true at t_1 , when Eric can force Fran to comply with her duty by making her bear a significantly greater cost. Thus, it's not clear that the pluralist account does very well at all in the *Rape* case.

So again, there's no victory here over internalism. As I argued above, an internalist account of liability to defensive harm cannot capture the case in which Fran inflicts harm that cannot avert the rape. But that's because *no* account of liability to defensive harm can capture the permissibility of inflicting these harms, because they're not defensive harms. And this claim seems to be pretty well borne out by Quong and Firth's attempt to capture them on such an account, which, as we've just seen, produces only an unconvincing result.

VII. Instrumentalism

So, I've argued above that internalism, properly understood, can withstand many of the criticisms that Quong and Firth level at it. But nothing I have said here will rescue

First draft: please do not quote or circulate.

internalism from what I identified at the outset as its central problems. Internalism still grants Murderer in *Lucky Escape* a permission to kill Victim if Victim decides to forcefully defend himself when he could have safely jumped, and Murderer cannot avoid the threat posed by Victim other than by killing him. This could entail that third parties may assist Murderer. On the internalist account, Murderer's moral status in this case is identical to that of a wholly innocent person whom Victim culpably attacks: Murderer is not liable to the harm, and so may use force to avert it. And, Murderer could have a claim to compensation if Victim succeeds in inflicting harm upon him.

These seem to me to be significant problems. Murderer is maliciously and deliberately trying to kill Victim. Unlike Quong and Firth, I do not think that if Victim kills Murderer in this case, he wrongs Murderer. One does not typically *owe it to one's attacker* to look for other ways to escape the threat that they pose. Murderer could not reasonably complain if Victim failed to do so ("but you should have run away, not defended yourself against me!"). Nor is Murderer like an innocent person whom Victim is culpably attacking. Thus, whilst it may be all-things-considered impermissible to unnecessarily harm a person who is culpable with respect to an unjust threat, it would not be impermissible because it wrongs that person. Indeed, as I suggest below, it's not clear that it is all-things-considered impermissible to unnecessarily harm a fully culpable aggressor.²⁶

I think we should reject internalism in favour of what I call an instrumentalist account of liability to defensive harm.²⁷ Whereas internalism holds a person liable to only the least harmful means of averting a threat, instrumentalism is less restrictive, holding a person liable

²⁶ We can think of other cases in which one might do something to a person that is wrong, but not because it wrongs them. For example, a doctor might assist the death of a patient at her request, but in doing so leave her children without financial support. The act of euthanasia could therefore be wrong, but not because it wrongs the patient.

²⁷ I don't compare externalism with instrumentalism because, essentially, I think that all externalist accounts have to be instrumentalist. That's the only way in which they can capture that these are accounts of liability to *defensive* harm.

First draft: please do not quote or circulate.

to instrumental harms – that is, harms that will avert the threat even if they are not the least harmful means. This liability is grounded in moral responsibility for an unjust threat.

Like internalism, instrumentalism is neutral between narrow and broad liability. Harms that aim at averting a threat other than the threat for which an aggressor is responsible are still instrumental harms. And, harms that aim at averting the particular threat for which the aggressor is responsible are also instrumental harms.

The difference between internalism and instrumentalism is illustrated by *Lucky Escape*. Internalism holds that since Victim can avoid the threat without harming Murderer, Murderer is not liable to be killed. Instrumentalism holds that since Victim can avoid the threat either by jumping or by killing Murderer, Murderer is liable to be killed.

What speaks against harming Murderer in *Lucky Escape* is the general moral requirement that one not inflict gratuitous harm. Members of the moral community in general can have a legitimate complaint against your inflicting such gratuitous harm, but the target of the harm does not have particular grounds for complaint. Since they are liable to the harm, harming them does not wrong them any more than it wrongs anyone else. This general moral requirement is indeed presupposed by the necessity condition. It's hard to see why defence would be governed by a necessity principle at all unless we think that there's a general moral requirement not to cause gratuitous harm. If we accept that defence is governed by a necessity constraint, as most people do, we have already granted the existence of a principle of the sort that I am suggesting.

(i) Less-than-culpable aggressors

But this principle as it stands isn't enough to capture an important feature of self-defence. Compare killing Murderer in *Lucky Escape* with killing minimally morally responsible threats like Resident in McMahan's *Evil Twin* case:

First draft: please do not quote or circulate.

Evil Twin: Victim breaks down in a remote area. He knocks on the door of an isolated farm to ask to use the phone. Unbeknown to him, his identical twin brother has committed a series of gruesome murders and is thought to be hiding out in just this area. Warnings have been issued to local residents. When Resident opens the door, she thinks Victim is the violent murderer come to kill her and tries to shoot him. Victim can shoot Resident before she shoots him.²⁸

Let's agree, for the sake of argument, that Resident is morally responsible for the unjust threat she poses to Victim.²⁹ Killing either Murderer or Resident unnecessarily is wrong on the externalist view. But killing Resident unnecessarily seems much worse than killing Murderer. If Victim can just jump off Resident's porch, getting out of her line of fire, it seems very bad indeed if he kills her instead (assuming that he realises that she's made a pretty reasonable mistake). Certainly, it seems much worse than killing Murderer, who is a fully culpable aggressor.³⁰

Quong and Firth's humanitarian duty is sensitive to the cost to Victim of refraining from killing a person. But it isn't sensitive to the worseness of killing Resident compared to killing Murderer, as long as the cost to Victim of refraining from killing either is the same. The cost one has to bear to fulfil one's humanitarian duty is a fixed cost, being grounded in "urgent need and not by appeal [...] to responsible choices."³¹ It is thus explicitly insensitive

²⁸ McMahan, 'The Basis of Moral Liability to Defensive Harm', *Ethics*

²⁹ There's probably some disagreement about this – it seems to me that she takes what Otsuka calls a 'moral risk' when she tries to kill Twin that's sufficient to ground responsibility – but those who reject responsibility in this case can substitute for some other case in which there seems to be only minimal moral responsibility for the unjust threat.

³⁰ Thanks to Jonathan Parry for first drawing my attention to this problem.

³¹ Quong and Firth, p; 693

First draft: please do not quote or circulate.

to a person's culpability or lack thereof. There's no requirement that Victim bear more of a cost to avoid killing a person who is only minimally morally responsible for posing a threat.

This means that, as well as being unable to capture the moral worseness of certain killings, there will be cases in which the pluralist account holds that killing Resident is permissible when (it seems to me) killing her is impermissible. Even if he'll break his leg if he jumps off the porch, Victim ought to do that rather than kill Resident. Assuming that a broken leg counts as a serious cost, Quong and Firth's humanitarian duty has no purchase in these cases since that duty requires that Victim bear only a low cost to fulfil it.³² Given this, they cannot deem killing Resident impermissible in this case.

If there's a moral requirement not to cause unnecessary harm, it's plausible to think that it's morally worse to cause unnecessary harm to people who have only minimal or moderate responsibility for posing a threat – who are, for example, partially excused because they act under duress or have made a reasonable mistake – than it is to cause unnecessary harm to culpable people. Such people are still liable to be killed in virtue of their moral responsibility for an unjust lethal threat. But the necessity requirement – that is, the requirement to use the least harmful means at one's disposal – is sensitive to an aggressor's responsibility. Whether or not a given means *counts as* the least harmful means depends in part on the aggressor's responsibility.

Most people think that Victim may discount harms to an aggressor: he need not treat them as equal to harms he can bear himself when deciding what is the least harmful means of averting a threat. Imagine the following case:

Choice: Attacker is going to break Victim's leg because he hates Victim.
 Victim can avert this by either (a) breaking Attacker's leg, avoiding all

³² Quong and Firth, p. 699

First draft: please do not quote or circulate.

harm to himself, or (b) breaking Attacker's arm, at the cost of also breaking his own wrist.

Most people do not think Victim violates the necessity condition if he chooses (a) and breaks Attacker's leg. Even though a broken wrist is objectively less harmful than a broken leg, (a) still counts as the least harmful means available because Victim can weight breaking his own wrist more heavily than breaking Attacker's leg. Harms to Attacker are heavily discounted when determining the least harmful means.

But it's plausible to think that this asymmetry between Victim and an aggressor weakens the less morally responsible the aggressor is. Faced with a minimally responsible threat, Victim must treat harms to her as nearly as weighty as harms to himself. Victim ought not to break Resident's leg rather than suffer a broken wrist. Nor should he kill her to avoid breaking his leg by jumping off the porch. Of course, working out how to weight and compare harms is always tricky, and is more than I can undertake here. But as an opening suggestion, perhaps the ratio of harm that Victim is required to bear in choosing alternative means that do not harm the aggressor is based on the amount of harm he would be required to bear if the aggressor were directly inflicting that harm upon him. It would, I think, be impermissible for Victim to kill Resident to stop her from breaking his leg. Thus, if he has to break his leg to avoid killing her, he's required to do so. But he'd be permitted to kill a fully culpable aggressor who threatened to break his leg, and thus he's not required to suffer a broken leg rather than kill Murderer.

The question, then, is whether allowing what counts as the least harmful means to differ between culpable and less-than-culpable aggressors entails that less-than-culpable aggressors are not *liable* to as much harm as culpable aggressors, precisely because greater harms are unnecessary. If so, it looks as if we've arrived at internalism. But I want to resist

First draft: please do not quote or circulate.

that – I think that proportionality operates in two ways in accounts of self-defence. The first is in determining what an aggressor is liable to: if she's morally responsible for a lethal unjust threat, then lethal force is proportionate. But proportionality also influences whether defence is necessary, because it influences whether force is the least harmful means available. It must do, because 'least harmful' is a moralised concept – it entails a judgement about how to weight harms to Victim and to an aggressor. Here, proportionality comes from a different direction and tells us what Victim is allowed to do, which may not equate to what an aggressor is liable to. That is, after all, the familiar externalist picture of the gap between liability and permissibility.

(ii) Counter-defence [This is the sketchiest part of the paper. Comments on this part are especially welcome!]

What does this mean for an aggressor's rights of counter-defence? On the one hand, my view holds that Murderer is liable to instrumental harm and thus he can't fight back. But on the other hand, it also looks like Victim fails in a duty if he inflicts unnecessary harm. This could be killing Murderer when he can jump at no cost, or killing him when he could jump at a low cost. This makes it sound like Murderer has at least some right of counter-defence, since he surely gets to do something to make Victim comply with his duty, just as a third party could.

But I argued above that in cases such as *Rape*, a culpable aggressor can be liable to harms other than defensive harm. I suggested that these other grounds must explain why Fran can break Eric's wrist even when this has no hope of averting the rape, since inflicting a harm that is not defensive cannot be justified by the target's liability to defensive harm. A fuller treatment of these issues will need to spell out these other grounds of liability to harm more clearly, but for now I want to pay attention to the fact that whatever these other grounds are – punishment, symbolism, assertions of dignity – they will also apply in cases in which

First draft: please do not quote or circulate.

Victim can avoid the harm from a culpable aggressor by, for example, using his parachute. Even if harming Murderer is not necessary on grounds of defence, it could still be permissible on these other grounds. If so, Victim will not fail in a duty if he harms Murderer. Thus, Murderer will not have a right to make Victim comply with a duty, and third parties will not be permitted to assist him in this. Nor will Murderer have a right to compensation. Whether this is the case will depend on what these other grounds are and how much harm we think they justify. [need to think more about the relationship between necessity and these other justifications for harming. Bit stuck here.]

But these other grounds for harming an aggressor become weaker as the culpability of the aggressor diminishes.³³ In the case of a minimally responsible threat like Resident, they may not exist at all. Then there's nothing to counter-balance or cancel out the fact that Victim uses unnecessary force if he harms her when he could, for example, jump off the porch at no cost to himself. Harming her in the absence of any other justification for using force would be impermissible. In this case, Resident and third parties can use force to make Victim comply with the general duty not to cause unnecessary harm, and that since this duty is more stringent with respect to less-than-culpable people, they might be able to use fairly significant force, especially given the argument I sketch above about imposing additional costs on those who initially fail to comply with their duties.

Conclusion

The instrumentalist account holds that liability to defensive harm can itself justify the infliction only of instrumental harms – that is, harms that are capable of averting a threat. Thus, whilst a person may be liable to more than the least harmful means of averting a threat,

³³ Worth noting that the argument for treating the baseline as that where luck only improves the threat's situation is weaker in the case of an MRT, because this threat isn't trying to bring it about. She still knew that this is what she was potentially exposing herself to – that she might cause a situation in which an innocent person can save his life only by killing her – but she's not trying to bring it about – she'd presumably be thoroughly relieved if V runs off or can escape without killing her. Need to think about the implications of this a bit more.

First draft: please do not quote or circulate.

she cannot be liable on grounds of defence to harm that is not a means of averting a threat at all.

I have argued that Jon Quong and Jo Firth's recent attempt to defend a non-instrumentalist account of liability fails. I have also argued that internalism is in much better shape than they realise. But I share with Quong and Firth the intuition that internalism is undermined by its granting to Murderer a permission of counter-defence akin to that of an innocent person, should Victim use unnecessary force against him. Instrumentalist allows that Murderer is liable to be killed by Victim, in virtue of his moral responsibility for an unjust threat to Victim's life. I have argued that the same is true of less-than-culpable, but still responsible, aggressors. What might nonetheless prohibit Victim's harming of these aggressors is the necessity constraint that applies to defensive killing. But whether a use of force counts as the least harmful means is sensitive to the aggressor's responsibility. Victim may heavily discount harms to a culpable aggressor. Combined with other possible justifications for harming, this may make it all-things-considered permissible for Victim to harm a fully culpable aggressor even when he need not do so to avert a threat to his life. But harms to minimally responsible aggressors are not heavily discounted, and alternative justifications for harming may not apply to such aggressors. In these cases, the aggressor may have a right of counter-defence if Victim attempts to use more than the least harmful means to avert a threat they pose.