

Self-Defense and Culpability: Fault Forfeits First

RICHARD J. ARNESON*

TABLE OF CONTENTS

I.	INTRODUCTION	231
II.	FAULT FORFEITS FIRST	232
	A. <i>Fault Forfeits First</i>	235
	B. <i>Accommodation</i>	237
	C. <i>Projectile Threat</i>	237
	D. <i>Guilty Bystander</i>	238
	E. <i>Guilty Victim</i>	239
III.	CRITICISM OF THE ACCOUNT WITH RESPONSES	240
IV.	THE CLAIMED IRRELEVANCE OF WHAT WE CANNOT KNOW	245
V.	CULPABILITY, WHAT?	246
VI.	CULPABILITY <i>WITH RESPECT TO A SITUATION</i>	251
VII.	NECESSITY	252
	A. <i>Unnecessary Self-Defense</i>	254
VIII.	PROPORTIONALITY, FUTILE SELF-DEFENSE, AND CULPABILITY	258
IX.	MORAL CULPABILITY AND MORAL PRAISEWORTHINESS	262
	A. <i>Hero</i>	262
	B. <i>Aggressive Heroine</i>	263

I. INTRODUCTION

Under what conditions is it morally permissible to kill someone in order to save your own life—or the life of another who is threatened? There

* © 2018 Richard J. Arneson. Professor of Philosophy, University of California, San Diego. The author thanks several helpful critics and commentators, with special thanks to Alec Walen.

seem to be clear cases. Threatened by an assailant who is trying to kill you for no good reason, you may use lethal force if necessary to save yourself from death or serious injury from the assailant's attack. Threatened with death in the form of an onrushing runaway truck, you may not save yourself by using a bystander or imposing on a bystander in a way that inflicts severe harm on her. In a justly celebrated essay, Judith Thomson notes that it is permissible to kill a "Villainous Aggressor" when necessary to save oneself from grievous harm, but adds that the evil intent — and thus culpability—of the Villainous Aggressor is not a necessary condition for justified self-defense.¹ She suggests that if we remove the elements of evil intent and culpability, and imagine an "Innocent Aggressor" whose actions pose a threat of causing one grave physical harm, one gets a similar moral permission to kill in self-defense.² Moreover, if we remove the element of agency altogether, and imagine an "Innocent Threat," whose bodily movements pose a threat of inflicting grave physical harm on one, but who is not at fault in any way for posing this threat, and is not doing anything at all, one gets a similar moral permission to kill in self-defense.³ Still, Thomson asserts, the differences between Villainous Aggressor, Innocent Aggressor, and Innocent Threat make "no moral difference"—to the permissibility of self-defense.⁴ Hence, according to Thomson, "it is permissible for you to proceed in Innocent Threat just as in Villainous Aggressor and Innocent Aggressor."⁵ Pondering these suggestions, she proposes, "I fancy we overrate the role of fault in many areas of moral theory."⁶

II. FAULT FORFEITS FIRST

To show that fault and culpability play more of a role in determining when it is permissible to kill to save one's life than Thomson allows, we simply need to vary the numbers of people who must be killed to save a threatened individual's life. To a first approximation, it is morally permissible to kill any number of Villainous Aggressors to save the life of just one threatened victim. Morality prefers the outcome in which the threatened victim lives even though this requires killing not just one but many Villainous Aggressors—even an entire army of them.⁷ I say "to a first

1. Judith Jarvis Thomson, *Self-Defense*, 20 PHIL. & PUB. AFF. 283, 283–84 (1991).

2. *Id.* at 284–85.

3. *Id.* at 287.

4. *Id.*

5. *Id.*

6. *Id.* at 286.

7. See Uwe Steinhoff, *Proportionality in Self-Defense*, 21 J. ETHICS, 263, 271 (2017) [hereinafter *Self-Defense*].

approximation” because this wide permission to kill Villainous Aggressors when necessary to save one’s own life gives out at some point. Would it be acceptable to kill a billion Villainous Aggressors to save one’s life? Surely not.

In contrast, even if we were to agree with the controversial judgment that it is justifiable to kill a single Innocent Aggressor to save one’s life, and with Thomson’s even more controversial judgment that it is morally justifiable to kill a single Innocent Threat to save one’s own life, these judgments would shift quickly if we imagine that more than one innocent person must be killed if one is to survive the threat.⁸ If morality permits us to kill virtually any number of Villainous Aggressors to save one’s life, but at most a few Innocent Aggressors, and perhaps at most a single Innocent Threat, a likely explanation and justification is that the culpability or nonculpability of the people whose behavior or bodily movements threaten to cause one harm make a large difference to what it is morally permissible to do in response to a threat to one’s physical safety.⁹

This Article explores the role that culpability should play in the ethics of justified killing with special attention to cases of self-defense. Someone who is culpable or blameworthy is an eligible recipient of blame for an action one has done, a choice one has made, a neglect or inattention one has shown, or a trait one possesses. A culpable person is a responsible agent who has failed to conform to moral requirements and lacks a complete

8. See Thomson, *supra* note 1, at 284, 287.

9. Uwe Steinhoff disagrees; he maintains morality permits one to kill any number of Innocent Aggressors if necessary to guarantee one avoids suffering grievous harm from their attack. See generally Uwe Steinhoff, *Just War Theory: Self-Defense, Necessity, and the Ethics of Armed Conflicts* 53, 71, 96 (2016) [hereinafter *Just War Theory*] (unpublished manuscript) (on file with author). However, I doubt his view results from his adoption of a Thomson-type perspective on the slight importance of culpability. More likely, he is moved by the thought that a person, whose right not to be attacked is being violated, has a broad license to resist the attack. In other contexts, he stresses the importance of culpability, holding, for example, that unnecessary attacks on morally innocent aggressors are impermissible whereas unnecessary attacks on culpable aggressors are permissible. *Id.* at 63–64. In this latter context, he cautions against “the modern ‘liberal’ philosophy professor’s warm-hearted, aggressor-friendly intuitions.” *Id.* at 61. Here he is defending the idea that the culpable aggressor forfeits his right to life, and is not wronged if he is killed, no matter what the reason. Steinhoff adds, “[t]he unjust aggressor has broken his ‘contract’ with you to respect each other’s rights, so you are under no obligation anymore to respect his rights.” *Id.* For my own aggressor-friendly intuitions, see Parts 6 and 7 of this Article.

excuse for the failure.¹⁰ Culpability varies by degree.¹¹ For purposes of this Article I shall suppose that what renders a person culpable or blameworthy is failure to show due consideration to the rights and interests of other people.

However, this formulation just stated might seem palpably inadequate. Surely one can be either blameworthy or blameless for failing to show due consideration to the rights and interests of other people. Perhaps one nonculpably has come to accept an incorrect account of what constitutes due consideration for others, so when one dutifully gives others due consideration according to this incorrect account, one is actually failing to show due consideration for others. But in these circumstances, this failure does not render one blameworthy. So I shall propose that we should interpret the due consideration requirement as subjective: what one fundamentally owes to others and to oneself is to make a conscientious effort to determine what treatment others should receive and what treatment one owes to them and how properly to balance consideration for others and concern for oneself—and then to make a good faith effort to conform one’s conduct to what one conscientiously takes to be what one owes them. Treating people with due consideration and concern, in the way that matters for culpability, is making a conscientious effort to figure out what conduct one owes others and striving conscientiously to conform one’s conduct to that standard.

But this subjective interpretation of due consideration as conscientious effort prompts questions and objections, some of which this essay tries to address in Part VI. This aspect of the account is more controversial than the generic idea of *fault forfeits first*, which could be accepted by readers who reject the subjective construal of culpability.¹²

In addition to holding that culpability can amplify one’s liability to defensive harm and lack of culpability can dampen such liability, I suggest a further role: culpability establishes a morally mandatory queue of eligible recipients of defensive violence. One must choose to direct one’s defensive violence only at the person or persons highest in this queue whose killing would save the victim from harm. These ideas are expressed in the “fault forfeits first” principle:

Fault forfeits first: In a situation in which the bodily movements of some place others under threat of suffering physical harm, resulting in a predicament in which someone must die, if one person is significantly morally at fault—culpable—with respect to this situation, and at fault—

10. See discussion *infra* Parts V–VI.

11. Jeff McMahan, *The Basis of Moral Liability to Defensive Killing*, 15 PHIL. ISSUES 386, 395 (2005) [hereinafter *Basis of Moral Liability*].

12. See discussion *infra* Parts V–VI.

culpable—to a significantly greater extent than others with respect to this situation, then it is morally preferable that this culpable agent should be the one who dies.¹³ If there is a choice of saving an innocent life by killing a single culpable rather than a nonculpable agent, one morally is permitted only to save the innocent by killing the culpable.¹⁴

Here is a broader formulation that extends the idea beyond self-defense:

A. *Fault Forfeits First*

All else being equal, the morally preferred target for violence directed at undoing or preventing a serious evil or injustice is an agent who is seriously culpable with respect to that evil or injustice and more culpable than anyone else involved in the situation.

Both formulations should be understood as assuming that the number of people who might be killed by different choices one contemplates would be the same. Therefore, fault forfeits first does not say one ought to kill any number of more culpable persons, however large, rather than kill a single less culpable person to avert an injustice. The numbers matter.¹⁵

More broadly, fault forfeits first is not a license to employ violence to avert wrongful harm in ways that would violate a properly formulated proportionality condition.¹⁶ If one can avert a rather small injustice only by imposing death or very serious harm on a seriously culpable agent, who is most eligible for being a target of violence according to fault forfeits first, the amount of harm one would have to inflict to avert the harm may be excessive in relation to the appropriately weighted wrongful harm one would thereby be averting.¹⁷ In this case proportionality is violated. Thus, attacking the culpable would be morally wrong. Moreover, a full statement of the proportionality condition would have to cover cases in which the likelihood and amount of wrongful harm that would ensue if one inflicts violence on the culpable vary and the comparative and absolute culpability

13. SHELLY KAGAN, *THE GEOMETRY OF DESERT* 24 (2012).

14. Richard J. Arneson, *Just Warfare Theory and Noncombatant Immunity*, 39 *CORNELL INT'L L.J.* 663, 668–69 (2006) [hereinafter *Just Warfare Theory*].

15. *See id.* at 670 n.17.

16. *See generally* Suzanne Uniacke, *Proportionality and Self-Defense*, 30 *L. & PHIL.* 253 (2010) (discussing standards of proportionality).

17. *See id.* at 256 (“Proportionate self-defense requires that the threat fended off/interest protected pass a threshold of comparable seriousness in relation to the harm inflicted on the attacker . . . harm that exceeds a threshold of comparable seriousness is impermissible because [it is] disproportionate.”).

of the target of violence deployed to prevent or mitigate wrongful harm also varies.¹⁸ A morally sensitive cost and benefit calculation is called for when the option of using violence to reduce injustice is being contemplated.

The import of fault forfeits first is to assert that culpability trumps the factor of causal involvement in generating wrongful harm. Tom may be acting in a way that is a major or sole cause of impending harm to someone and Tomasina may be culpable with respect to this situation but not causally involved in posing a threat of harm to anyone; yet, fault forfeits first can single out Tomasina as the morally mandatory target of any violence that might be deployed to reduce wrongful harm.

As stated, fault forfeits first does not take a stand for or against the moral permissibility of killing an Innocent Aggressor or Innocent Threat if one faces a stark choice: allow the threatened victim to die or kill an Innocent Aggressor or an Innocent Threat.¹⁹ Fault forfeits first does not involve any commitment to the claim that moral fault is a necessary condition for justified killing in self-defense.²⁰

However, if two perfectly blameless individuals have the misfortune of interacting in a way that threatens them with the imposition of death or severe injury, I suggest that morality does not favor attaching the harm to one of these persons rather than the other. If one is blamelessly attacking the other, principles attuned to culpability should hold the fair distribution of harm is that—given one must suffer harm—each should be accorded an equal chance of avoiding it.²¹ This does not entirely deny a right of self-defense to the wrongfully attacked victim; there is surely no more moral aptness that the harm should fall on him rather than on his attacker. Ideally a coin flip should determine who suffers the unavoidable harm.

Here are some examples to illustrate the idea.²²

18. See Just War Theory, *supra* note 9, at 19.

19. See Heidi M. Hurd, *Justification and Excuse, Wrongdoing and Culpability*, 74 NOTRE DAME L. REV. 1551, 1572 n.39 (1999).

20. *Just Warfare Theory*, *supra* note 14, at 668–69.

21. *Id.* at 672.

22. To reiterate, this Article makes several suggestions that are differentially controversial. One suggestion is that culpability is a significant factor that affects whom it is permissible to attack to mitigate unjust threats. Another is that fault—culpability—forfeits first. A third is that one being culpable with respect to the situation is a necessary condition for being permissibly attacked. A fourth is that culpability to any degree does not open the door to being permissibly attacked to mitigate an unjust threat unless attacking someone is necessary to this end.

B. Accommodation

You are being chased by a villainous aggressor who is trying to kill you and will succeed, unless you jump up into a niche where the aggressor cannot follow. There is a bystander standing in the niche who has a perfect right to be there. The bystander sees that if he accommodates you by moving to the back of the niche, you can jump to it and both of you will be safe. If the bystander does not move in this way, you could still save yourself, but only by jumping to the niche and jostling this bystander, causing a fatal fall. The bystander also sees that this is the case. The bystander declines to move to the back of the niche.²³

In this case I suppose it is morally permissible to jump to the niche, killing the bystander and saving your life. Given the circumstances, the bystander's failure to be reasonably accommodating renders him, compared to you, significantly and culpably at fault.²⁴

C. Projectile Threat

Suppose you are standing at the bottom of a cliff. Someone falls onto you from above. As it happens, the trajectory of the individual's fall will bring it about that he will land on you, killing you, and because he happens to be well padded he will live no matter where he lands. As it happens, you are unable to move out of the way to escape the body that is hurtling toward you, but you do have the option of using your ray gun to vaporize the person whose body is threatening you. In that case he dies and you live.²⁵ This is an example of the Innocent Threat case that has elicited opposed responses from Judith Thomson, Jeff McMahan, Michael Otsuka, and others.²⁶

Fault forfeits first says that if someone in this situation is seriously at fault with respect to it, and more culpable than anyone else involved in the situation, then that person is the person who ought to die if someone must die. Suppose the person falling on you is a rock climber who was

23. *Just Warfare Theory*, *supra* note 14, at 669.

24. See e.g., HELEN FROWE, DEFENSIVE KILLING 21–45 (2014) (providing various examples in which one person is located where his body impedes another's escape from wrongful attack and discussing each example with a view to revising the distinction between a bystander and a threat).

25. See ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 34 (1974).

26. There is a considerable amount of literature on this controversial issue. See, e.g., Michael Otsuka, *Killing the Innocent in Self-Defense*, 23 PHIL. & PUB. AFF. 74, 74–94 (1994); Thomson, *supra* note 1, at 287; see also JEFF MCMAHAN, KILLING IN WAR 32–37 (2009) [hereinafter KILLING IN WAR] (discussing different interpretations of innocent).

climbing without a rope, even though she knew her route would take her just over a much traveled hiking path at the crux move, and even though her skill level is not so outstanding as to make falling at that point a very unlikely event. The person is then reckless with respect to the risk she imposes on you, and fault forfeits first yields the plausible implication that self-defensive violence—if necessary to save the life of the threatened victim—is morally acceptable.²⁷

On the other hand, suppose you have put yourself in harm's way on purpose, for no legitimate reason. The area is one where sublime but very tricky and dangerous rock climbs are located, and falls not that uncommon. Only the most experienced climbers attempt these climbs without a rope but, attracted to the sublimity of the climbs, and the added aesthetic beauty of climbing without a rope, many of the very best climbers flock to this area and—reasonably and justifiably—attempt these climbs, and falls do occur. You have passed by warning signs that say “Danger! Climbing area ahead: Falls common. Stay well away from the bottoms of the cliffs!” Moved by the idle and mildly sadistic curiosity that impels fans of fights in hockey games and spectators at car races to hope for fights and fiery crashes, you position yourself at the bottom of the cliff to have the best view of a fall should one occur. You have no thought that you might act as a cushion for a falling climber and anyway the chances of this occurrence are small. However, as it happens, someone does fall on you, as described above. In this situation you are culpably reckless with respect to the situation that has unfolded and—given your culpability—you may not save your life by vaporizing the nonculpable person who falls toward you.²⁸ Again, fault forfeits first yields plausible implications for this sort of case.

D. Guilty Bystander

You are driving up a narrow mountain road and encounter an Innocent Aggressor driving downhill in an armored vehicle that will soon smash your vehicle.²⁹ (In an alternative scenario, let the threat barreling down on you be an innocent threat.) Your only lifesaving alternatives are to blow up the ongoing threat or to turn your vehicle onto the shoulder of the road. The shoulder is occupied by a person who just happens to be there and has a right to occupy the shoulder, but is wrongfully and culpably desirous of your death, hopes the threat you face materializes, and is doing his best to harm you by throwing snowballs at your vehicle to distract your

27. See *Just Warfare Theory*, *supra* note 14.

28. See *id.*

29. See *id.* at 671.

attention. This person's actions are entirely ineffectual and pose no threat of harm to you.

In this case, I suppose one is morally required not to save your life by killing the Innocent Aggressor who threatens you, given that there is an alternative way of saving your life that would result in the death of the guilty bystander, a significantly culpable agent—one who is significantly more culpable than anyone else involved in this incident. This is so, even though this bystander bears no causal responsibility for the threat to your life that you face and is behaving in a way that is actually harmless.³⁰

If fault forfeits first is correct, then we should be able to imagine that anyone involved in a self-defense scenario might be the person who ought to die if anyone does—even the victim of the threatened or developing assault. Blaming the victim may sometimes be appropriate because the person who is most blameworthy with respect to a threatened or developing assault might, in some circumstances, be the intended victim of this assault.³¹ Consider the following example:

E. Guilty Victim

Suppose I have been maliciously tormenting my wife with false accusations of sexual infidelity on her part. Coming from me, these accusations are deeply wounding and inflict terrible suffering on her. Finally, she reacts by throwing a punch at me. She intends to hurt me physically but not seriously wound me. Unfortunately, unbeknownst to her, I have just acquired a serious medical affliction, such that if her blow lands on my skull, it will

30. In the example in the text, the guilty bystanders are acting, though without effect, with intent to harm. In principle, fault forfeits first could identify—as the morally preferred person to die when someone's behavior or movement threatens another with serious harm—a guilty bystander who is ill-disposed to the victim and directs hateful thoughts at her, and who would harm the victim wrongfully if he could. However, this guilty bystander has no action available and hence is neither acting nor trying to act toward the victim at all. This concept embraces the suggestion in *Just Warfare Theory*, *supra* note 14, at 671. *But see generally* VICTOR TADROS, *Causation, Culpability, and Liability*, in *THE ETHICS OF SELF-DEFENSE* 110 (Christian Coons & Michael Weber eds., 2016) (endorsing a narrower version of fault forfeits first).

31. This situation has been described as the “Provocateur Limitation.” Joshua D. Brooks, *Deadly-Force Self-Defense and the Problem of the Silent, Subtle Provocateur*, 24 *CORNELL J.L. & PUB. POL'Y* 533, 540–44 (2015) (providing an example of when a victim is more culpable in an altercation).

certainly kill me.³² Her assault on me is unjustified and morally wrong but, given the circumstances, just barely morally wrong, and she is just barely culpable for launching it. In contrast, my despicable behavior toward her is unexcused and extremely blameworthy. I cannot duck or otherwise evade her approaching blow, but I can draw a weapon and kill her in an instant, thereby preventing her blow from landing on me with lethal result.

However, because she is hardly culpable at all for her assault and I am very culpable with respect to this situation, fault forfeits first singles me out as the person who ought to die in this situation if anyone does. I would be morally permitted to evade or duck the blow that threatens me if that were possible, but being a culpable victim, I lack a self-defense right to save my life by killing the person whose assault threatens to kill me, even though defending myself by acting with lethal force against my attacker is the only way that I could survive.³³ My significant culpability with respect to this situation brings it about that, to this degree, I have forfeited my moral right to self-defense.

III. CRITICISM OF THE ACCOUNT WITH RESPONSES

(A) Judith Thomson resists the idea that we should draw the line of moral permissibility between Villainous Aggressor and Innocent Aggressor, in which case one may kill to save one's life—or the life of a threatened victim—only if the person one kills is a wrongful Aggressor who is culpable for acting wrongly.³⁴ Let us concede that the Villainous Aggressor deserves punishment for his aggression. Thomson comments, “But who are you, private person that you are, to be dishing out punishment to the villainous for the things that they do?”³⁵

In response: It is false that one who accepts fault forfeits first must suppose private persons are morally entitled to inflict punishment in the self-defense scenarios depicted. Punishment is the deliberate imposition of suffering on an offender for his offense and is appropriate only when the punishing agent has authority to impose such suffering and only when the person who is punished has been found guilty of an offense that merits punishment by means of a fair procedure for determining the accused's guilt or innocence. We may safely suppose these conditions on justified punishment are not met when a person who is at risk of being a victim of

32. See Steve P. Calanrillo & Dustin E. Buehler, *Eggshell Economics: A Revolutionary Approach to the Eggshell Plaintiff Rule*, 74 OHIO ST. L.J. 375, 377 (2013) (discussing liability for defendants who harm victims with unknown conditions or unusual responses).

33. See Brooks, *supra* note 31, at 544–46.

34. Thomson, *supra* note 1, at 284–85.

35. *Id.* at 285.

culpable aggression must decide what to do.³⁶ Let it be granted that when a reasonably just government has been established, private persons standardly lack the authority to carry out punishment for aggressive violence—or at the least, they lack authority to impose severe sanctions including imprisonment and death.³⁷ However, in the situations governed by fault forfeits first, someone must die, and the question is, *who is the most morally appropriate person to be singled out for death?* Whether one is entitled to punish is irrelevant to settling the issue.

Further response: It is anyway conceivably morally appropriate for private persons living under a government to act in a punishing capacity.³⁸ Sometimes the government lacks capacity for enforcement.³⁹ For example, a seriously culpable wrongdoer is making his escape and afterward will never be within reach of the law. Suppose you know all this, and in particular are in an excellent epistemic position to know for sure that the offender is guilty and deserving of at least some quantum of punishment. Here your epistemic command substitutes for carrying out a fair procedure to establish guilt or innocence; in this temporary state of nature situation the government's unique entitlement to punish arguably lapses, and we can suppose your infliction of punishment on the wrongdoer will do some good by incapacitation or deterrence. If your carrying out punishment would not create a scandal and erode respect for the law, punishing the offender might be acceptable in some circumstances. But to reiterate, the claim that one is bound to select a culpable offender for harm if it is the only way to prevent harm to others in a situation does not rely on any claimed authority to inflict punishment on the offender.

(B) Recent philosophical discussions of culpability usually treat it differently from what this essay proposes. The extant accounts regard it as an extra constraint on justified infliction of self-defensive harm.⁴⁰ Roughly, the

36. See Thomson, *supra* note 1, at 286.

37. See R.A. Duff, *Political Retributivism and Legal Moralism*, 1 VA. J. CRIM. L. 179, 182 (2012).

38. See *id.* at 181.

39. See *id.* at 180.

40. Jeff McMahan considers—without accepting—culpability in this way. JEFF MCMAHAN, *THE ETHICS OF KILLING: PROBLEMS AT THE MARGINS OF LIFE* 403 (2002) [hereinafter *ETHICS OF KILLING*]; *Basis of Moral Liability*, *supra* note 11, at 389–405; see also Kimberly Kessler Ferzan, *Justifying Self-Defense*, 24 L. & PHIL. 711, 734 (2005) (accepting culpability as a necessary condition for becoming liable to a defensive harm). *But see* Jeff McMahan, *Self-Defense and Culpability*, 24 L. & PHIL. 751, 763–764 (2005) [hereinafter *Self-Defense and Culpability*] (discussing culpability not associated with moral responsibility).

idea is that to be morally permitted to kill another to save oneself from death or grievous harm, it must be the case that one has a right not to be subject to the threat of harm one faces. The person one would kill to save oneself from this harm is morally responsible for imposing this risk of harm, and, in addition, this person is culpable for imposing this risk. Against such an account, the objection is that culpability is not a necessary condition for justified infliction of self-defensive harm.⁴¹ Either the first two conditions just listed suffice for a justification or some other substitute set of conditions not including a culpability requirement suffices.

To introduce a complex type of case, suppose an unjust war of aggression is in progress and being resisted by warriors with a just cause.⁴² The soldiers fighting on the unjust side of the conflict are misinformed about the relevant facts and believe they are waging a just war. Furthermore, they are in no way culpable for being misinformed in this way. When these morally innocent soldiers attack people—who in fact have a right to not be harmed—it is, to put it mildly, highly implausible to hold that the soldiers' complete lack of culpability renders them morally ineligible targets of violence necessary to thwart their, in fact, unjust actions.⁴³

As stated, fault forfeits first does not insist moral culpability is a necessary condition for justified infliction of self-defensive harm, so the objection to the deployment of a culpability condition described above does not pose any objection to this norm.⁴⁴ But this may seem to be a case where the theorist is out of the frying pan and into the fire. Fault forfeits first allows that when somebody is considering a violent act against another to save his own life, the culpability of the other, if sufficiently great, and sufficiently greater than that of others who are involved in the situation and might instead be subjected to attack—may by itself suffice to justify the exercise of lethal force.⁴⁵ This means that when caught up in a situation in which someone's behavior or bodily movement is threatening to cause harm to others, a person may be legitimately liable to being attacked even though she is neither doing anything nor behaving in such a way—nor is her body

for imposing or causing an unjustified harm or threat of harm and holding that moral responsibility, absent culpability for a sufficiently serious wrongful harm, can suffice for liability to defensive violence).

41. Richard Arneson, *Resolving the Responsibility of Dilemma*, in *THE ETHICS OF WAR* 67, 81 (Saba Bazargan-Forward & Samuel C. Rickless eds., 2017) [hereinafter *Resolving the Responsibility*].

42. *See id.* at 73.

43. Seth Lazar presses this issue to argue for weaknesses in Jeff McMahan's revisionary just war doctrine. *See* Seth Lazar, *The Responsibility Dilemma for Killing in War: A Review Essay*, 38 *PHIL. & PUB. AFF.* 180, 181 (2010); *see also Resolving the Responsibility*, *supra* note 41 (stressing culpability in contrast to McMahan's position).

44. *See supra* pp. 232–235.

45. *See Resolving the Responsibility*, *supra* note 41, at 81–82.

moving in such a way—that poses any threat of harm to any other person.⁴⁶ This feature of fault forfeits first may strike some as obviously morally outrageous.

In response: We should not be stampeded by this concern into abandoning fault forfeits first. We should simply recall the examples already considered, especially Accommodation and Guilty Bystander, and consider whether the judgments proposed for these situations are acceptable after reflective scrutiny.⁴⁷ The underlying idea behind these judgments is that the moral shield that insulates a person from being the permissible target of a harmful attack can be forfeited in whole or in part. What triggers forfeiture of moral immunity from being subject to attack is egregious failure to show due consideration for others' legitimate moral rights and interests.⁴⁸ Failure to have and show such concern tends to threaten others with receipt of wrongful harm in many situations, but the failure has moral significance in and of itself. When an individual fails to show due concern for others, the reciprocal character of the obligation of constraints dictates that certain behaviors toward that individual that would otherwise be required by the due concern requirement are, to some degree, lifted or suspended.⁴⁹

Of course, in many settings, sheer ill will toward others does not trigger a right for them to impose otherwise unacceptable harm on the person manifesting this ill will.⁵⁰ If I stick pins in a voodoo doll made to resemble my neighbor, thereby intending to inflict bodily harm on him by voodoo magic, there is actually no risk of harm generated for anyone; therefore, my neighbor is not warranted to exercise force imposing harm on me. However, in unusual circumstances, where somebody's behavior or bodily movement does impose risk of harm on a person who has not forfeited any of her rights not to be harmed, and where one may save the victim from harm by violent act, then fault forfeits first becomes applicable. In these circumstances, naked culpability not associated with any act that wrongfully threatens harm to others—and unassociated with any causal involvement in the harm that defensive violence aims to avert—can justify imposition

46. *Id.*

47. *See supra* pp. 237, 238–39.

48. *See* Just War Theory, *supra* note 9, at 21.

49. *See* David Rodin, *The Problem with Prevention*, in *PREEMPTION: MILITARY ACTION AND MORAL JUSTIFICATION* 143, 164 (Henry Shue & David Rodin eds., 2007) (“[M]any rights are implicitly reciprocal [and] on a plausible understanding of rights, one only has the right to life so long as one respects the right to life of others.”).

50. *See* NOMY ARPALY, *UNPRINCIPLED VIRTUE: AN INQUIRY INTO MORAL AGENCY* 172–73 (2003).

of harm against the culpable when necessary to protect innocent victims from harm.

Fault forfeits first claims neither that culpability is necessary nor sufficient for justified self-defensive killing. The claim is rather that culpability is a significant factor in determining whom, if anyone, it is acceptable to kill in these scenarios, and in particular that culpability under certain conditions establishes an ordering of morally preferred candidates for being the targets of violence when violence is necessary to save an innocent person's life.

(C) In law, and perhaps in morality, culpability in the air, unconnected to wrongful action, does not suffice for forfeiture of the culpable person's rights, including rights against being attacked.⁵¹ Jonathan Quong writes, "Individuals can be blameworthy for acting on bad motives, but so long as they do not violate anyone's rights, their moral rights are unaffected."⁵²

In a similar spirit, Seth Lazar writes,

Suppose that you can save a saint from a threat only by killing a thief; the thief had nothing to do with the threat at all, but he has stolen in the past. The thief's bad character and past crimes are irrelevant to whether he is liable to be killed. He is not connected to this threat, so he cannot be liable to be killed to avert it.⁵³

In response: The law does not and should not perfectly mirror the content of fundamental moral principles.⁵⁴ The law, with its associated social norms, should be designed to function as an effective and fair means to promote conformity of people's behavior to what fundamental morality dictates given the empirical conditions we all face, including epistemic and motivational gaps and deficits. In most actual self-defense scenarios, agents will have no way of calculating who is most culpable in the situation and will be tempted to twist their spontaneous assessments to suit their interests. So the law might rightly downplay the importance of culpability for liability to defensive harm in familiar ways whereas fundamental moral principles do not.⁵⁵

However, at the level of fundamental moral principle, we can agree that bad character irrelevant to the situation at hand does not render one liable to permissible violence aimed at correcting or averting unjust wrongs. But being blameworthy with respect to the situation at hand is a different matter and, I submit, can trigger liability, even without any exercise of agency that is causally connected to wrongful harm or threat of such harm.

51. See Jonathan Quong, *Liability to Defensive Harm*, 40 PHIL. & PUB. AFF 45, 54 (2012).

52. *Id.*

53. SETH LAZAR, SPARING CIVILIANS 131 (2015).

54. Richard Arneson, *Discrimination, Disparate Impact, and Theories of Justice*, in PHILOSOPHICAL FOUNDATIONS OF DISCRIMINATION LAW 87, 95 (Deborah Hellman & Sophia Moreau eds., 2013).

55. See Quong, *supra* note 51, at 45, 47.

Suppose my thieving tendencies express themselves in the situation in which the saint is in peril, but not in a way that causally produces injustice or threat of injustice. Or suppose the most culpable agents in a war situation are the soldiers fighting for the just cause.⁵⁶ They are acting from sheer race hatred; the fact that their acts are not violations of any moral rules is an accidental fluke. If one can advance the just war cause equally well by killing some enemy soldiers—who are completely blameless for fighting for a bad cause—or one’s culpable comrades, fault forfeits first says one should shoot the culpable fellow soldiers rather than the blameless enemy.

IV. THE CLAIMED IRRELEVANCE OF WHAT WE CANNOT KNOW

Someone concerned with formulating moral rules that provide good practical guidance to agents who are wondering if it is acceptable to attack others to save one’s life and legal rules that facilitate the criminal justice system’s functioning to enforce these practical moral rules will balk at this Article’s reliance on improbable thought experiments. We are seeking guidance for agents in situations they will likely face, not solutions to hypothetical puzzles lacking any relevance to real-world decision problems.

Hence, one might object that in canonical self-defense scenarios agents will never be in a good or even a passable epistemic position to make the assessments of moral praiseworthiness and blameworthiness required to determine what fault forfeits first implies regarding what is permissible conduct in their actual decision problem. The discussions in this Article simply distract us from the task of formulating moral rules for guidance.

In response, we make moral judgments for evaluative purposes as well as for practical guidance. Studying the behavior of some distant historical figures, we might—from intellectual and moral curiosity—wish to determine whether what they did was morally acceptable. Someone killed another to save her own life—was this morally permissible? We might have a moral and intellectual interest in answering similar questions regarding current proposed actions—even if no issue of practical guidance accompanies the determination. If Smith is going to employ violence to save himself from injury even if morally impermissible, the question remains, is Smith’s plan morally permissible? Even if we can never see the inner workings of someone’s soul, Smith’s behavior may provide us valuable evidence regarding his moral blameworthiness or praiseworthiness.

56. *See supra* p. 242.

In some cases, a person wondering whether attacking another is permissible might be in a prime epistemic position to judge the culpability of those involved in the situation. Whether these cases are rare or frequent is irrelevant. We can imagine an Innocent Aggressor scenario in which the person who becomes the threatened victim of the aggression has been deliberating for days with his back-fence neighbor, who is troubled about what to do in a situation he fears he might face. His conscientious efforts to determine what is right and his serious resolve to do what he thinks is right strikes the interlocutor as exemplary. To his surprise, the interlocutor finds himself a victim of aggression and sees the aggressor is his back-fence neighbor, surely an innocent, fully nonculpable aggressor. By some similar coincidence, the Innocent Aggressor's threatened victim might also have good evidence concerning the serious culpability of the nearby bystanders—whom he could kill to save his life.

We should distinguish (1) what should be the laws governing permissible uses of violence in self-defense or defense of others along with the social norms and publicly proclaimed morality governing these matters from (2) what are the fundamental moral principles that hold universally and timelessly and determine what in any situation ought to be done.⁵⁷ Laws, social norms, and public morality should be selected with the aim of promoting adherence by people to fundamental moral norms. Some factors that bear on the issue of what morality permits, requires, and forbids at the fundamental level, should perhaps disappear or appear only in simplified form at the secondary and tertiary levels, because their retention would be counterproductive in terms of encouraging people to adhere as closely as possible to the pattern of conduct that fundamental principles enjoin.

V. CULPABILITY, WHAT?

What is culpability? To some extent I hope my defense of fault forfeits first stands independently of the question, *what conception of culpability or moral blameworthiness is best?* Readers might agree with my claims about fault forfeits first but disagree with my understanding of culpability, or readers may agree with me on culpability but not on fault forfeits first. However, whether fault forfeits first is plausible will depend on how culpability is understood.

Consider some sample views:

57. See generally Just War Theory, *supra* note 9.

Nomy Arpaly holds that a person's moral worth can be positive or negative.⁵⁸ Negative moral worth is moral blameworthiness.⁵⁹ Having moral worth, or lacking it, depends on one's good or ill will.⁶⁰ Arpaly writes,

If good will—the motive(s) from which praiseworthy actions stem—is responsiveness to moral reasons, deficiency in good will is insufficient responsiveness to moral reasons, obliviousness or indifference to morally relevant factors, and ill will is responsiveness to sinister reasons—reasons for which it is never moral to act, reasons that, in their essence, conflict with morality.⁶¹

Also, a person can be more or less indifferent to moral reasons and more or less powerfully motivated by the desire to act on sinister reasons.⁶² According to Arpaly, whether one desires to act on sinister reasons depends on whether the reasons really are sinister—not on whether one believes them to be sinister.⁶³ The same holds for one's obliviousness or indifference to moral reasons.⁶⁴

Writing specifically about the morality of self-defense, Larry Alexander defines a “culpable aggressor” as one who meets two conditions. First, the person “intends to commit an act in the future that he believes imposes risks of various magnitudes of harm on various types of [victims].”⁶⁵ Second, the person “believes that that the circumstances that will probably exist at the time he acts, discounted by his estimates of the probabilities of their existence,” will be thus and so, and their being thus and so in fact renders the act he intends to commit unjustifiable—whether or not he believes his act, as it will occur in the circumstances as he supposes them to be, will be unjustifiable.⁶⁶

A generic conception of what it means for an agent to be fully culpable for a wrong act is that the agent was responsible for doing that act and lacked an excuse of false belief or duress. According to both Arpaly and Alexander, falsely believing that one's act is justified by a certain moral principle does not qualify as a good excuse if the postulated moral principle

58. ARPALY, *supra* note 50, at 69.

59. *Id.*

60. *Id.* at 79.

61. *Id.*

62. *See id.* at 101.

63. *See id.* at 80.

64. *Id.* at 79.

65. Larry Alexander, *Recipe for a Theory of Self-Defense*, in *THE ETHICS OF SELF-DEFENSE* 20, 21–22 (Christian Coons & Michael Weber eds., 2016).

66. *Id.*

is false.⁶⁷ In this sense false moral beliefs do not remove or diminish culpability.

These views, along with other common views of culpability, share the feature that one can be culpable for doing something wrong even if one lacked a fair opportunity to have done the right thing instead. For example, I might kill Sam under circumstances that render the act morally wrong, but—through no fault of my own—I did not realize the act was morally wrong because—through no fault of my own—I have come to embrace false moral principles. I might be motivated by sinister reasons, reasons from which it is never moral to act, reasons that in their essence conflict with morality.⁶⁸ For example, suppose utilitarianism is wrong and some opposed doctrine of moral rights is correct. Doing what maximizes utility is then never an adequate ground for action because any utility-maximizing act might also conflict with some person’s moral right, which ought to be honored. In the example, assume killing Sam violates his moral right and is morally impermissible on that basis. If the empirical facts are as I suppose them to be, my act of killing Sam is unjustifiable, so my act is culpable according to Alexander’s account as well.⁶⁹

“Through no fault of my own” needs to be parsed. Suppose I worked hard to figure out the correct moral principles—seeking advice, deliberating hard, discerning what others who might serve as authority figures or helpful interlocutors believe, checking out their credentials and the quality of their arguments, deliberating further, and so on. However, I am inept at moral deliberation and I go badly astray. Of course, it is my fault that my conscience is now wrongly calibrated. My practical reasoning is defective. This is a fault in me. However, this may be due to factors entirely beyond my power to control—bad early socialization, poor genetic endowment, and so on—so in another sense, my errors are not really my fault.

Arpaly’s view captures an important ideal of moral praiseworthiness.⁷⁰ Someone whose actions are responsive to correct moral reasons and who is strongly motivated to bring it about that his acts are responsive to what are in fact correct moral reasons is surely morally praiseworthy.⁷¹ But the antonym of being morally praiseworthy in this sense is not being morally blameworthy but rather something like being worthy of moral dispraise.

Whether our choices and actions lie within our power to control is of course a profoundly contested issue. In this Article I pursue some implications of supposing that the necessary conditions for being morally praiseworthy

67. See ARPALY, *supra* note 50, at 104; Alexander, *supra* note 65, at 22.

68. See ARPALY, *supra* note 50, at 109–10.

69. See Alexander, *supra* note 65, at 22.

70. See generally ARPALY, *supra* note 50.

71. *Id.*

or blameworthy—deserving or undeserving—set forth in the control principle are sometimes satisfied. I do not argue this supposition is true. I further believe if we reject hard determinism and accept a view that leaves room for individual responsibility, we should still accept it would be unfair to hold individuals responsible for what lies entirely beyond their power to control. As noted above, arriving at correct beliefs about what morality requires may be beyond one's power to control.

A conception that may perhaps meet the stated constraint will have to be subjective. Whether one is morally praiseworthy—in the morally blameworthy sense—depends not on whether one acts correctly for the right reasons but on whether one makes a conscientious, good faith effort to do so. This will include making a conscientious, good faith effort to recognize correct principles that determine what is morally right.

The rough idea here is that—in some form or another—the question, *what do we owe to one another or how should we treat others*—including other sentient creatures—is bound to occur to us. One becomes morally blameworthy by failing to make conscientious efforts to learn the answer or for failing to make conscientious efforts to conform one's conduct to the requirements of morality as one perceives them. One becomes morally blameworthy by failing to make conscientious efforts to do the right thing. There are degrees of blameworthiness.

This account cannot be right as it stands. In some situations, making a conscientious effort to do what is right may seem inappropriate. For example, if one knows he will fail to do the right thing no matter how prodigious his conscientious effort is, then expending the effort is just wasteful. So, one becomes morally blameworthy by failing to make conscientious efforts to conduct oneself as morality dictates insofar as one deems such effort to be appropriate in the circumstances—with a further proviso that one must make conscientious efforts to see to it that one's appropriateness judgments are correct. There is a regress here I suppose to be harmless.

This account still cannot be right as it stands. My aim is to describe a conception of culpability such that it will not be a matter of sheer luck beyond your power to control whether you become culpable or not. But pursuing this aim may be a futile task, like a dog chasing its tail. For one thing, it will be a matter of luck whether you encounter choice situations in your life that elicit a response from you that renders you culpable. One person faces the stresses and temptations of a corrupt and charming boss, another does not. For another thing, it is a matter of luck beyond your power to control whether you happen to have the desire to put forth a conscientious

effort to show due consideration for others. Some will have the desire and some will not. Among those who do, the desire will be a powerful motivation for some but not others.

A response to the latter worry is that we must adjust a person's raw deservingness score that registers the degree to which she puts forth conscientious effort as appropriate by allowing for the fact that the ensemble of circumstances people face, including their given pro-moral desires, render it harder for some and easier for others to put forth conscientious effort. However, any being with rational agency capacities that meet the threshold level of personhood will have some curiosity about what one owes to others and some concern to give what one owes.

A response to the former worry is that it can indeed be a matter of sheer luck that one person's good or evil dispositions face situations in which they manifest in choice and action while another person's do not. Whether one is culpable with respect to a particular situation is thus partly beyond one's control, insofar as it is beyond one's power to control what decision problems one faces and what situations one confronts. Of course, what decision problems one faces may lie within one's power to control to a degree. One should avoid the near occasions of sin. But this *should* is pro tanto; one's conscientious judgment might be that morality requires one to enter the lions' den even if this means he will then face a severe temptation to be wrongfully mean toward the lions.

Suppose two persons dispose themselves to do evil and are strictly culpable for doing so. If the two persons are equally firmly disposed to do evil and—by sheer luck—one faces a situation that triggers the disposition and the other does not, they are equally morally culpable. Sheer luck beyond their power to control cannot render one more culpable than the other. But then we must recognize that people's culpability may not be fully comparable. Suppose three persons are equally culpable and equally, but not fully, disposed to do evil. Two of the persons face a temptation and one does not, and of the two tempted persons, one by dint of hard effort resists the temptation and the other, who could have done exactly the same, does not, and does evil. There may be no determinate fact of the matter as to whether the person who by luck evaded the temptation would have succumbed if the person had not evaded it, and so no determinate fact of the matter as to whether the one who escaped temptation is more or less morally deserving than the one who faced it and succumbed. But the no-moral-luck constraint does not govern any and all moral judgments we might make, just those concerning a person's moral worth, moral praiseworthiness, blameworthiness, or basic deservingness score.

VI. CULPABILITY WITH RESPECT TO A SITUATION

The idea of culpability that figures in this Article is being culpable *with respect to a situation*. The italicized phrase is crucial, though vague. As Seth Lazar notes, merely being a generally bad person or having bad moral character would not suffice to render one liable to self-defensive harm.⁷² If you are under threat from an approaching freight train and could save your life only by killing a bystander, the fact that the bystander is a bank robber or cheats outrageously in high stakes card games, does not render the bystander a legitimate target of attack.⁷³ For that, the bystander's culpability must be relevant to the situation at hand.⁷⁴ The bystander must be involved in the situation in some seriously blameworthy way, or else her culpability does not trigger any forfeiture of her moral immunity from being attacked.

Noticing culpability in the air—entirely unrelated to the threat the mitigation of which justifies violence—is irrelevant to liability to self-defensive harm, does nothing to rule out culpability unconnected to causation of wrongful threat as a determiner of who may be attacked.⁷⁵ For example, a grievously wrongful intention or disposition of will can do the job. In the freight train example, suppose the bystander is here and now firmly disposed to inflict wrongful harm on the person threatened by the train if he could do so. He is not acting on the disposition because he sees no way to attack the threatened person, so he is not posing any wrongful threat. I submit that the bystander who is culpable in this way has rendered himself a permissible target of violence if that will mitigate the threat to the innocent person.

Jeff McMahan describes an example in which you are under lethal threat from two culpable assailants acting entirely independently.⁷⁶ It is a matter of coincidence that they and you are all in a dark alley at the same time, which turns out to be a good venue for assault. Your body is exposed to one attacker and his shots will soon kill you, but you have no effective means of defense.⁷⁷ However, you can shoot the other attacker, from whose attack you are shielded, and you see that if you do, his falling body will

72. See LAZAR, *supra* note 53, at 131.

73. See *id.*

74. See *id.* at 131–32.

75. See *id.* at 127–28, 132.

76. *Self-Defense and Culpability*, *supra* note 40, at 757 (describing what McMahan calls the “Basement Window” case).

77. *Id.*

obstruct the genuinely threatening attack being launched by the other.⁷⁸ McMahan acknowledges he is inclined to judge that you would be justified by attacking the nonthreatening culpable attacker to save yourself from harm.⁷⁹ But, he worries there is no principled way to distinguish this case from others in which a wrongful threat could be averted by attacking a culpable individual or one of bad moral character where the culpability intuitively has nothing at all to do with the threat being averted, but where by some coincidence killing the culpable would bring about reduction or elimination of this threat.⁸⁰

McMahan is right to worry. It is desirable to have a principled norm to guide intuitive judgment. This we lack. But it would be a mistake to be inhibited on this basis from recognizing that where, as here, the agent's serious culpability arises from some feature of his agency that involves morally inappropriate attitudes toward the threat that the threatened person faces, he may become a permissible target of attack. In McMahan's example, we suppose the ineffective attacker is disposed to welcome the attack by the other and to assist that attack if he could. We need to add that we must also suppose the disposition does not simply afflict the agent in ways beyond his power to control, but arises in him and is sustained by processes for which he should be deemed accountable. To the degree this is only partially so, his culpability diminishes; if it is not so at all, culpability vanishes.⁸¹ In another example, we might imagine an onlooker who celebrates the impending harm about to fall on the innocent person and is made gleeful by the prospect of grievous harm to this innocent. Here too there are seriously morally wrong attitudes directed at what is happening to the threatened person in this situation.

VII. NECESSITY

Consider the moral requirement that self-defense, to be justified, must be necessary to avert or mitigate the harms that would result from the wrongful attack that is claimed to be the legitimate object of self-defense. As Seth Lazar and others have noticed, the necessity idea invoked here is not transparent.⁸² Suppose that a threatened victim could choose any of a

78. *Id.*

79. *Id.* at 758–59.

80. *Id.* at 764–65; *Basis of Moral Liability*, *supra* note 11, at 386, 393.

81. For a clear statement of the position that the control principle is never satisfied, so the conditions for holding agents to be morally culpable or praiseworthy are never obtained, see 1 DEREK PARFIT, *ON WHAT MATTERS* 264–66, 268 (Samuel Scheffler ed., 2011).

82. See FROWE, *supra* note 24, at 88–119; Joanna Mary Firth & Jonathan Quong, *Necessity, Liability, and Defensive Harm*, 31 L. & PHIL. 673, 674–75 (2012); Seth Lazar,

range of defensive actions against an aggressor, the actions differing in their associated probabilities of lessening to varying degrees the threatened harm to the victim and in their associated probabilities of causing harm of varying magnitudes to the aggressor.⁸³ Where along these scales does a proposed self-defense fail to satisfy the necessity condition? This Article sets this important issue to the side.

Here I make the vague suggestion that the victim ought to shoulder some costs rather than impose harm as great as death on an attacker. If the threatened victim can with certainty escape a wrongful attack by parachuting to safety, at a cost of a sprained ankle, or could alternatively save herself from harm by shooting her attacker dead, she must parachute. Also, the victim must accept some risk of grievous harm to herself rather than shoot to kill or the equivalent. If parachuting to safety would give her a .999 chance of saving herself and a .001 chance of death, she must take the escape option. The intuitive idea is that an attacker's well-being interests are discounted morally, but not discounted to zero, by virtue of his wrongful action.⁸⁴

Also, I suggest culpability, if any, on the part of the attacker lessens the weight of his welfare interests in the necessity calculation and for that matter the victim's culpability, if any, triggers a similar discounting of her welfare interests. A culpable victim must shoulder greater costs and greater risk of harm rather than shift all risk and costs to attacker if she can and the greater the culpability of the attacker, the less solicitude for his well-being is morally dictated. In this determination, a comparative weighing of the victim and attacker's culpability is called for.

There is a limit to this discounting. If the victim has an available course of action that eliminates all risk that she will suffer serious harm from the wrongful attack that is in the offing, the victim is morally required to take that option rather than impose unnecessary harm on her attacker.⁸⁵ If this issue is correctly framed in terms of forfeiture of rights, the point is that no matter how culpable the attacker would be, the attacker does not forfeit the right to minimal consideration of the sort that is in play when one refrains from imposing death or grievous injury on a person when there is no

Necessity in Self-Defense and War, 40 PHIL. & PUB. AFF. 3, 4 (2012) [hereinafter *Necessity in Self-Defense*].

83. For a discussion of this issue, see *Necessity in Self-Defense*, *supra* note 82, at 9–21.

84. *See id.* at 21–23.

85. *Id.* at 15.

significant cost to oneself attached to the refraining. Hitler merits minimal consideration. Every person does.

We should note a further issue, a loose end left dangling. What benefits, in terms of wrongful harm diminution or mitigation, can count in the determination of whether proposed self-defense is necessary? Suppose one can save oneself from aggressive attack by parachuting to safety, but one has available an alternative action that will kill his aggressor and thereby (a) save oneself from harm and also (b) prevent him from certainly—or with high probability—aggressing against others in the near future. Is attacking the aggressor in this situation to secure the extra harm avoidance in (b) an act that satisfies the necessity condition? We could also imagine a variant case in which, by counter-attack, the victim could secure (b) but not (a).

Whether or not one believes the counter-measures to wrongful attack described above to be self-defense, one should allow that if one is permitted to kill an aggressor in given circumstances when necessary to save one's own life, one should also be permitted to kill an aggressor if necessary to prevent that person from wrongfully killing another. Moreover, if attacking an attacker to save another is permissible, it is surely permissible if—along with saving another one also saves oneself—one gets (a) and (b) gains together.

Culpability also plays a role in untangling rights and wrongs, and permissible and impermissible action, in cases in which the necessity condition for justified self-defense is breached and a victim of threatened attack wrongly counter-attacks when she could certainly escape all harm by availing herself of an escape route. Culpability here figures in the determination of what should be done via fault forfeits first. Consider a variant of an example introduced by Helen Frowe⁸⁶ for another purpose:

A. *Unnecessary Self-Defense*

Culpable Aggressor is attacking the victim, who has the options of escaping the attack by leaping to safety or instead shooting Aggressor.⁸⁷ She commences shooting. For anyone who adheres to a necessity condition for justified self-defense—as I believe we all should do—Frowe's victim is acting impermissibly. In these circumstances she lacks a right to attack her attacker. Does that imply that the moral tables are now turned, and Culpable Aggressor has a self-defense right to attack his intended victim? This implication might seem dubious.

First, let us suppose Culpable Aggressor is not a leopard who somehow changes his spots. Culpable Aggressor still has murderous designs on the

86. FROWE, *supra* note 24, at 88 (detailing the “Lucky Escape” example).

87. *See id.*

victim, even after the victim counter-attacks. Here fault forfeits first yields plausible verdicts. Suppose that the aggressor is just barely culpable for his aggression—perhaps his responsibility is diminished, though not extinguished, by innocent drug-taking, as would be the case if someone surreptitiously poured a mild aggression potion in his coffee. Suppose also that the victim's culpability is great—perhaps the aggressor is of Asian origin and the victim hates Asians and has been waiting for a chance to get away with killing one. Her serious and much much greater blameworthiness makes her the person who should die if anyone must die in this scenario. If the culpability comparison instead issues in the verdict that the victim is far less blameworthy than Culpable Aggressor, then all things considered he and not she is the eligible target of violence. If the fault forfeits first test does not yield a clear verdict, then it is wrong for each to be attacking the other and there is no morally preferred, much less required target of violence.

A variant is worth mentioning. Suppose just as the victim comes to the place where she must choose either to take the escape option or start shooting if she wants to live. Culpable Aggressor suddenly repents and either stops aggressing or becomes disposed to stop if he could—perhaps he has unleashed a robot attacker and he cannot now control it, though shooting him would somehow cause the robot to malfunction. So, one might say a new situation has emerged, in which the victim is now attacking either a former or unwilling aggressor. One might argue that according to fault forfeits first in the new situation, the victim is seriously culpable and the more seriously culpable member of the group of two of which one must die. Once again, the example raises the question of how to individuate *situations* for the purpose of deciding who is culpable and to what degree with respect to this situation. I think being responsible by way of past culpable aggression for the victim's present plight still renders one seriously culpable with respect to the current situation. So, it is not obvious that Culpable Aggressor may permissibly shoot at the victim in self-defense, even though it is taken for granted that the victim, in attacking, is acting impermissibly. But, the example reminds us that we need more of an account of being culpable with regard to the situation at hand than I am providing.

Finally, consider the related issue posed by stand-your-ground scenarios.⁸⁸ A Culpable Aggressor takes something by force that rightfully belongs to a the victim, or threatens to harm the victim unless the victim relinquishes some right. The victim could safely retreat, and no vital interest of his would then be at stake. Some hold it is morally permissible to respond with defensive violence, even lethal violence if need be, rather than yield—for the present—some small right to the Culpable Aggressor.⁸⁹ Such violence is not necessary to defend any vital interest or large right of the victim and seems disproportionate if executed as a means to protect the victim’s small right or nonvital interest.⁹⁰

Stand-your-ground scenarios differ in morally significant ways. Here is an argument that appears to vindicate the claim that it is morally acceptable to kill culpable aggressors to defend against small wrongful losses in one range of cases. Suppose A wrongfully harms B, perhaps taking what belongs to B. The taking does not harm B to such an extent as to justify killing A in response and reclaiming what belongs to B by right. But consider a series of proportionate responses. B forcefully acts to reclaim her rightful property, imposing a no more than proportionate harm on A. A responds with greater wrongful force. In response to this further wrongful attack, B escalates her defensive response, which is again no more than is proportionate to the cumulative harm and threat of harm she has suffered. A responds again with escalated force, and the cycle continues until B responds with lethal force. At this point, we are supposing that B’s lethal defensive response is no more than proportionate to the cumulative wrongful harms and threats of harm delivered by A.

Moreover, suppose that B correctly foresees this sequence will unfold if she responds to the violation of her rights with a series of proportionate responses. Suppose she instead attacks A preemptively, reasoning that if she engages in the step by step response, she will be undergoing some risk that A, at any point, might jump the queue and impose lethal violence on her. For her own safety, B attacks A with lethal force just after the initial

88. In stand-your-ground scenarios, a victim of a threat may use force, including possibly lethal force, to stand her ground against an aggressor “even if the victim could avoid [doing so by] retreating.” *No-Retreat Rule*, BLACK’S LAW DICTIONARY (10th ed. 2014).

89. For example, Heidi Hurd stipulates that stand-your-ground lethal violence may not be deployed in *de minimis* matters but allows that defending property or the right to remain standing where one has a right to stand in the face of a culpable aggressor’s demand to retreat can pass the threshold of rights that may be defended with lethal violence if there is no other way to secure the right here and now. Heidi M. Hurd, *Stand Your Ground*, in THE ETHICS OF SELF-DEFENSE 254, 254–56 (Christian Coons & Michael Weber eds., 2016).

90. See Seth Lazar, *National Defence, Self-Defence, and the Problem of Political Aggression*, in THE MORALITY OF DEFENSIVE WAR 22–32 (Cécile Fabre & Seth Lazar eds., 2004).

transgression by A. If lethal violence by B against A is justified in the step by step escalation, is lethal violence by B against A not justified in the version if the story in which she preempts the certain-to-occur sequence and responds with lethal violence against A at the first step?

Despite the plausibility of the arguments just stated, we should find them fallacious. Explaining why this is so is a bit tricky. Consider another case. A wrongfully insults B, for example, by asserting, “Your mother dresses you funny.” B foresees that if she responds, A will certainly respond with entirely unjustified lethal violence, and B will then have a right of self-defense to kill A to save her life in the face of this wrongful attack. In this situation, we should hold that B is morally bound to refrain from taking—what in other circumstances would be—a proportionate defensive response to A’s wrongful insult. She has a humanitarian duty toward A not to do what she knows would stir A to wrongful action leading to his death.

Morality can require forbearance, including forbearance in the face of violations of one’s moral rights. If A steals a few dollars from B, the innocent victim B is not permitted to kill A to reclaim her small bit of property. I suggest an extension of this idea. In this situation, B is not morally permitted to initiate steps that she knows will provoke disproportionate or flatly wrongful responses by A, leading to a situation in which B becomes liable to lethal defensive violence. This result would be reached by a consequentialist calculation employing any sensible estimation of the values at stake. More controversially, I submit that a sensible deontology, a morality of do’s and don’ts and of constraints and options, should reach a similar result in this sort of case.

There is a potential confounding factor here that I simply want to set aside despite its practical importance in likely scenarios. If A would respond with lethal force to a verbal counter-insult, he appears to be a loose cannon, likely to be dangerous in future interactions with whomever he encounters. Harming him, and even provoking him to bring about a situation in which he becomes liable to be killed, might be acceptable all things considered, taking these risks of future harms he might well perpetrate into account. Let us focus on the simpler case in which these complications are not present.

To see the point at issue, it may help to focus on situations in which only slight harms are at stake. Sometimes one’s associates and colleagues and friends behave in ways that are wrongfully inconsiderate, but still, the morally required response is that one overlook the slight infraction and carry on in good spirits, for the sake of harmony. A defensive response may provoke unjustified counter response and a hostile sequence that should

be avoided. The issue here is not long-term deterrence and long-term maintenance of peace, but simply avoidance of short-term bads. Part of one's duty to be kind to interaction partners is a duty to assist them to avoid harmful wrongdoing that will injure themselves and others. One may have this duty not only to people to whom one has special ties but also to strangers and acquaintances. Exactly the same is true when the stakes are raised and avoidance of lethal defense is in play.

The judgment that it is wrong to kill even a very culpable aggressor when doing so is not necessary to secure one's own vital interests or important rights, and would not bring about wider benefits, rests on a duty of humanity each of us owes to any person in any decision problem: do not seriously harm a person when nothing important is thereby gained. In making a moralized cost and benefit calculation that determines whether imposing harm on a person is justified, all things considered, the vital interests of the person one might harm count for a lot, even if those interests are somewhat discounted by the that person's culpability. "Somewhat discounted" indicates that being culpable never obliterates the moral value of bringing about gains or avoiding harms for the culpable.⁹¹ The same duty of care and concern for everyone and anyone can rule out doing what one would otherwise have a perfect right to do, except that one foresees that doing it would trigger wrongdoing by another that would make killing him a proportionate response. Culpability has its limits.

VIII. PROPORTIONALITY, FUTILE SELF-DEFENSE, AND CULPABILITY

Justified self-defense satisfies a proportionality condition.⁹² I have a right of self-defense against someone who wrongfully verbally assaults me, perhaps. I am licensed to defend myself by verbal retort and perhaps even by a physical shove or push—think of cases in which I lack the mental capacity for firing an effective verbal retort. But, it is morally impermissible for me to counterattack by launching serious physical blows or shooting the person in the leg. These acts would violate proportionality. The good to be gained by self-defensive action must be proportionate to the harm that is threatened.⁹³ A proportionate defensive harm is at least not greatly in excess of the harm that is averted.⁹⁴

Many commentators have noted that proportionality so understood includes a contestable "success condition."⁹⁵ In the evidential sense of justification,

91. See *Basis of Moral Liability*, *supra* note 11, at 395.

92. See Uniacke, *supra* note 16, at 269.

93. *Id.*

94. See Just War Theory, *supra* note 9, at 19.

95. *Id.* at 75–76.

one's self-defensive act, on the basis of the evidence available to the agent, must reasonably be expected to do enough good, by reduction of the harm threatened by the attack, in comparison to the harm threatened by the attack. If not, proportionality is violated. But, if the proposed self-defensive act launched by or on behalf of the victim does nothing to stop or mitigate the attack, this proportionality condition cannot be satisfied. This implication delivered by the many views of self-defense that accept the proportionality condition has seemed, to some commentators, to be a devastating bullet for these doctrines to bite.⁹⁶ Surely, they say, it is permissible for a rape victim to fight back against her attacker even if such attack will do no good whatsoever: will not reduce the harm she suffers and may aggravate it, will not deter the aggressor or other aggressors from inflicting wrongs on other potential victims, and will not bring about any other good either.⁹⁷

Appeal to the claims this Article defends—that culpability significantly affects the limits of permissible self-defense and also that fault forfeits first—do not resolve the puzzle here because proportionality is claimed to be a necessary condition on morally permissible self-defense. The intuitively plausible claim that entirely futile resistance to culpable aggression can be morally justified challenges proportionality as standardly understood. As Uwe Steinhoff observed, postulating that the Culpable Aggressor forfeits her right not to be harmed in some reciprocal way also fails to resolve the puzzle because the fact that the Culpable Aggressor lacks a claim right not to be harmed does not mean harming such an individual is morally justified.⁹⁸

In the face of the puzzle, some observers postulate some further good that seemingly futile self-defense might achieve, so proportionality is after all satisfiable in such cases.⁹⁹ Along this line, Daniel Statman proposes that wrongful aggressors challenge our status as persons with dignity and

96. Uwe Steinhoff is one of these commentators. His solution is to deny that it is a conceptual necessity that self-defense must aim to reduce the harm threatened by an unjust attack. *See* Just War Theory, *supra* note 9, at 9.

97. Steinhoff posits that (1) self-defense might aim at resisting attack without being accompanied by any aim of reducing its harm; (2) self-defense does not require a goal of success—one can strive to stay an aggressor's knife hand while knowing this attempt at resistance will fail; and (3) there simply is a basic right of self-defense—understood as a right of the victim and others to resist wrongful attack—that is not conditional on an expectation of gain from the resistance—at least when the victim is facing a culpable aggressor. *Id.* at 20.

98. *Id.* at 22.

99. *Id.* at 19–20.

worth; violent counter-violence gains the counter-attacker honor in the face of this challenge.¹⁰⁰ One wonders, however, why honor could not be regained by a calm assertion, in thought or speech, that the aggressor is wrong and that one is a person with equal dignity and worth, and why reclaiming honor in that way—if honor is ever at stake in such proceedings—is not the morally required stance.

Taken strictly, this challenge to the proportionality condition on permissible self-defense is off the topic of this Article, which is concerned with delineating the role of culpability in determining the boundaries of permissible and impermissible self-defense. However, we should accept a framing of these issues that presupposes that proportionality as standardly understood is acceptable. The challenge of futile self-defense makes me see that a more accurate formulation of the underlying fault forfeits first thought than we have been considering so far is that *if and only if someone must die when unjust wrongs are being perpetrated, then provided that any choice to kill targets the same number of individuals, it is morally required that the seriously culpable with respect to this situation, if more culpable than others whose deaths would serve the purpose, be chosen as the targets of violence.*

The unjust harm reduction project must target the especially culpable independent of their causal role in bringing about the unjust harm. This formulation allows that if no especially culpable agents are available whose deaths would serve this unjust harm reduction project, others may perhaps be targeted for violence for the greater good. The formulation leaves open what should be done when one can achieve a reduction of unjust harm by attacking larger number of especially culpable agents rather than a smaller number of individuals who might be deemed eligible on other grounds, but I have added that the greater the culpability of the especially culpable, the greater the case for killing them, rather than smaller numbers of others, to achieve the same good end. But, the “if and only if” phrasing rules out futile self-defense as morally impermissible.

The justification I offer appeals to the simple idea that every single person’s life is valuable, sacred if you will, regardless of the uses one makes of the rational agency capacities that confer personhood status. Whatever one has done, however morally bad one’s character, it is morally better to keep one alive and give one the opportunity to live well in the future and perhaps do good for others as well as oneself—at least when the opportunity provision is costless for others. If one can prevent injustice without harming even Hitler, he is morally required to do so, unless refraining from harming Hitler somehow imposes costs on others that they are not morally required to bear.

100. Daniel Statman, *On the Success Condition for Legitimate Self-Defense*, 118 ETHICS 659, 668 (2008); see also FROWE, *supra* note 24, at 123–60 (endorsing Statman’s view of this point). *But see* Just War Theory, *supra* note 9 (demurring).

In another terminology, rights forfeiture is always limited and conditional—never absolutely sweeping and across the board.

To this I simply add that the supposition that harming the culpable aggressor is completely futile is a strong assumption. Usually imposing costs on an aggressor has at least some slight tendency to deter future similar acts by this aggressor or others. Usually resistance to aggression has some chance of reducing the harm of the attack, even if the overall expected benefit is small.

Also, I note that even if we were to accept a retributivist view of punishment, in some cases, what looks like futile resistance, serves the cause of retributive justice in a rough and ready way. The victim may foresee that there is some chance, perhaps even a certainty, that the culpable aggressor will not be apprehended and brought to a fair trial and punished—the infliction of the harm of punishment on the culpable offender being in itself morally desirable according to retributivism. Because she is arguably in a good epistemic position to know the aggressor menacing her is truly culpable, she might justifiably engage in vigilante action that imposes on the culpable aggressor some of the punishment that is morally desirable for its own sake that he receive.¹⁰¹ In this way seemingly futile self-defense when inspected more closely will be seen sometimes to do some good and to satisfy proportionality.

This Article rests on a view of culpability that rules out retributivism as unjustified. Punishing a person for culpable wrongdoing must have expected positive value and expectably do some good for future potential crime victims, others, or the agent herself. The inherent value of each life is nonforfeitable and so the life of the most culpable agent has moral value and should not be sacrificed unless the sacrifice does some good. But, I do not claim to defend the rejection of retributivism presupposed in this Article. I simply note that according to this Article there are strict limits to the efficacy of culpability in justifying counter-violence. Hence, proposed self-defense that is futile, and so really fails to satisfy proportionality, is forbidden. But someone might accept some of the thinking about culpability asserted in this Article while also defending a retributivist view of punishment, and for someone whose thinking runs along this line, the puzzle of the

101. Judith Thomson explicitly rejects this vigilante punishment rationale for killing a culpable aggressor to save one's own life when it is threatened by his aggression. *See* Thomson, *supra* note 1, at 302. I suggest we should accept vigilantism at least in principle but reject retributivism.

seemingly moral acceptability of futile self-defense can in some cases be eased. What seems futile might not, under closer scrutiny, really look to be so.

IX. MORAL CULPABILITY AND MORAL PRAISEWORTHINESS

If culpability plays a significant role in determining who is an eligible target of violence when innocent persons are under threat, as argued here, one might wonder whether positive deservingness with respect to a situation can insulate one from such eligibility. To explore this issue, consider a variant of an example introduced by Jeff McMahan to illustrate his proposal that someone might be liable to defensive violence via being morally responsible for a threat even though one's conduct is morally faultless.¹⁰² He imagines a driver of a car, proceeding with due caution and having kept the car in good order by due diligence, who through no fault of her own "veer[s] out of control."¹⁰³ Driver's car skids toward a pedestrian walking on the adjacent sidewalk.¹⁰⁴ McMahan suggests the driver, being morally responsible by initiating a reasonable but risky action, is liable to self-defensive violence if necessary to save the pedestrian's life.¹⁰⁵ I am myself skeptical of this position. But consider the following scenario:

A. *Hero*

An ambulance driver heroically volunteers to drive through a minefield to deliver desperately needed medicine to wounded individuals. Returning through the same minefield, his ambulance skids out of control, despite his faultless, careful driving, and the vehicle threatens to crash into a pedestrian with lethal force.

My intuition is that even if—*big if*—one should agree with McMahan that in his example the faultless driver exposes herself to self-defensive violence, one should still demur at this same judgment regarding Hero. This suggests positive desert affects liability just as negative desert—blameworthiness. One might suspect our intuition about the case is clouded by the surmise that its being known that heroes are liable to be killed in such scenarios dampens the motivation of potential future heroes, with bad consequences. We stipulate this in this example no such consequences are in the offing.

102. See KILLING IN WAR, *supra* note 26, at 165 (detailing the "Conscientious Driver" scenario).

103. *Id.*

104. *Id.*

105. *Id.*

Consider another example—a variant on an Innocent Aggressor case:

B. Aggressive Heroine

An individual by dint of extraordinary virtuous attempts to discover the relevant facts of an ominous situation comes to have the false belief Smith is a culpable aggressor whose seemingly innocent action of walking across a park will trigger an attack that will kill many innocent people. The individual also believes that if she tries to stop Smith from unleashing this attack by running over him with her car, it is overwhelmingly likely Smith will kill her. Nonetheless she goes after him with her car. Smith actually is entirely innocent and is simply walking across a park, threatening no one. He happens somehow to be apprised of the situation. He can save his life only by shooting the heroic aggressor.

I submit that we ought to hold that Aggressive Heroine is insulated from liability to self-defensive violence in this situation by her great deservingness in this situation. Even if—*big if*—we should hold that a threatened person would be permitted to kill an innocent—but not especially positively deserving—aggressor if doing so was necessary to save his life from the threat posed by the aggressor’s conduct, the judgment should shift as we vary the example by cranking up the innocent aggressor’s positive deservingness.

However, accepting that fault forfeits first has a companion principle in praiseworthiness forfeits last does not require any commitment to symmetry. Maybe culpability matters more, in rendering an individual a morally apt target of self-defensive violence, than moral praiseworthiness matters in rendering an individual an inapt recipient of such violence.

