

Defense and Desert: When Reasons Don't Share

KIMBERLY KESSLER FERZAN*

Assume Culpable Aggressor threatens Innocent Victim with a knife. Victim is stronger than Culpable Aggressor and is able to defend herself by punching Culpable Aggressor in the face, causing him to stumble back and drop the knife. Not only was this action necessary, but also Victim believed it to be so to save her life.

I take it that this is an uncontroversial case of self-defense. My question is whether this is also a case of punishment. Uwe Steinhoff suggests that it might be.¹ Indeed, he states that “nothing hinders an act from being *both* punitive and defensive. In fact, this double structure is probably the normal one for cases of self-defense.”²

Now consider a different case. Assume Victim is weaker than Culpable Aggressor. Culpable Aggressor is threatening to slap Victim in the face. Because of the size differential, the only harm Victim can inflict that will be sufficient to stop the attack is to shoot Culpable Aggressor in the arm. Although I will discuss this case further below,³ let us assume for now that

* © 2018 Kimberly Kessler Ferzan. Harrison Robertson Professor of Law, Caddell and Chapman Professor of Law, University of Virginia. I thank participants in the Workshop on Self-Defense at the University of San Diego as well as the UVA incubator lunch for helpful discussion of this paper. Special thanks are due to Rachel Harmon, Debbie Hellman, Doug Husak, Adam Kolber, Patrick Tomlin, Uwe Steinhoff, Alec Walen, and Peter Westen, as well as to my research assistant, Isabel Tuz.

1. Uwe Steinhoff, *Just War Theory: Self-Defense, Necessity, and the Ethics of Armed Conflicts* 11–12 (2016) (unpublished manuscript) (on file with author).

2. *Id.*

3. *See infra* pp. 274–77.

shooting someone in the arm is disproportionate to the threat posed by slapping someone in the face. However, as it turns out, Culpable Aggressor previously broke someone's finger and due to a procedural error, will never be punished by the state. May Victim now shoot Culpable Aggressor in the arm if the shooting is proportionate to the combination of proportionate defensive force and deserved harm?

Here, we are—or at least I am—tempted to think Victim may combine these justifications and shoot Culpable Aggressor. In contrast, Steinhoff finds this idea “ill-conceived” and “odd.”⁴

I argue below that Steinhoff is wrong on both counts. Most cases of self-defense are not instances of punishment. And, there are rare cases in which adding self-defense and punishment can justify inflicting more harm. Ultimately, I will suggest that when desert and defense are both sufficient reasons to justify both the rights forfeiture and the infliction of harm, both reasons are not typically simultaneously operative. And I will argue that these reasons can aggregate when they are both necessary and only jointly sufficient to justify the rights forfeiture or harm imposition. Finally, I will address how we should determine, in individually sufficient cases, whether the action—if intended as both—is desert or defense.

Let us start with some preliminaries about self-defense. Jeff McMahan claims that “[t]o attack someone who is liable to be attacked is neither to violate nor to infringe that person's right, for the person's being liable to attack *just is* his having *forfeited* his right not to be attacked, in the circumstances.”⁵

McMahan takes this forfeiture to be “highly specific. . . . [A]ll that is necessary is the forfeiture of the right not to be attacked *for certain reasons*, by certain persons, in certain conditions.”⁶ Liability is thus a nuanced form of forfeiture. It is also *instrumental*:

[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, deter him (or perhaps others) from future wrongdoing, or compensate a victim of his prior wrongdoing. The goal is internal to liability, in the sense that there is no liability except in relation to some goal that can be achieved by harming a person.⁷

Hence, to McMahan, what it means to be liable to defensive force is for the aggressor (1) to have forfeited rights and (2) for there to be a defensive reason

4. Steinhoff, *supra* note 1, at 13.

5. JEFF McMAHAN, *KILLING IN WAR* 10 (2009).

6. *Id.*

7. *Id.* at 9.

to impose the harm.⁸ According to McMahan, without a defensive reason, there cannot be liability to defensive force.⁹

McMahan offers us a liability *package*, and I think it is important to understand the motivation for this combination of factors. First, McMahan sees himself as asking a question within “preventive justice,” where the idea is that there is a harm to be distributed, and the question is simply upon whom the harm should fall.¹⁰ Second, McMahan is after an account that itself yields the conclusion that the defender’s act is justified.¹¹

But this formulation of the inquiry becomes uncomfortably stilted in two respects. First, this way of asking the question—as distributing an unavoidable harm—puts the rabbit in the hat. The idea that we have a *goal* of distributing *inevitable harm* assumes conceptual connections between instrumentality and liability that need an argument, not a stipulation.¹² Second, even if we are stipulating, and thus cabining our inquiry, it is hard not to slip into an analysis that equates liability with forfeiture alone. Indeed, at points, McMahan himself appears to fall victim to this slippage. In *Targeted Killing: Murder, Combat or Law Enforcement?*, McMahan investigates the possible justificatory grounds for the killing of Osama bin Laden.¹³ McMahan asks what would happen if someone killed bin Laden and bin Laden was not actually planning another attack.¹⁴ McMahan’s answer is that appearances can count.¹⁵ After noting that one may “forfeit[] any claim to the benefit of the doubt,” McMahan further claims that he may not fight back and that, “[h]e is, one might say, liable to be killed on the basis of a mistake that he is responsible for making it reasonable for others to make, even though he is not liable to be killed for defensive reasons.”¹⁶

8. *Id.* at 10.

9. *Id.*

10. Jeff McMahan, *The Basis of Moral Liability to Defensive Killing*, 15 PHIL. ISSUES 386, 395 (2005) [hereinafter *Basis of Moral Liability*]. I often think of this, roughly, as tort law with a time machine—go back and distribute the harm to the person who should have borne it in the first place.

11. *Id.* at 386.

12. See Steinhoff, *supra* note 1, at 48–49.

13. See generally Jeff McMahan, *Targeted Killing: Murder, Combat or Law Enforcement?*, in TARGETED KILLINGS: LAW AND MORALITY IN AN ASYMMETRICAL WORLD 135 (Claire Finkelstein, Jens Davis Ohlin & Andrew Altman eds., 2012).

14. *Id.* at 138.

15. *Id.* at 138–41.

16. *Id.* at 140.

I submit that the only way to render that claim intelligible is to divorce liability from instrumentality. To be sure, one might forfeit rights to certain epistemic or procedural protections, but that still would not explain why it is actually instrumentally good to kill someone who does not pose a threat. Make no mistake about it: McMahan's view of necessity is fact-relative. On McMahan's view an individual is not liable to be killed just because a mistaken defender reasonably believes he is.¹⁷ Indeed, McMahan has argued that mistaken defenders, who reasonably believe they are under attack, are themselves the liable parties.¹⁸ But from the objective, fact-relative perspective, there is no reason to kill an apparent attacker. Even if it is the victim's fault the defender makes a mistake, mistakes are not reasons. There may be evidence-relative reasons and belief-relative reasons, but there are no fact-relative reasons to act on a mistake, and McMahan's definition of liability is fact-relative.¹⁹ This means that what McMahan is really after here is forfeiture.

In other work, I have speculated that liability language arose because of fear of "forfeiture."²⁰ Forfeiture seemed too permanent, too theoretically problematic, too inapt to account for the suspension of the aggressor's rights during the attack.²¹ Enter liability. But liability has itself become a cluttered and contested concept.

17. See Jeff McMahan, *Just Cause for War*, ETHICS & INT'L AFF., Dec. 2005, at 1, 8 n.11 (2005) ("I cannot be made liable by your mistake, even if it is a reasonable one.").

18. See McMahan, *Basis of Moral Liability*, *supra* note 10, at 389. That being said, I think there is a deep equivocation in McMahan's views about the importance of evidence-relative beliefs. McMahan sometimes slips into probability talk that is not consistent with his otherwise fact-relative views. See, e.g., Jeff McMahan, *Who is Morally Liable to Be Killed in War*, 71 ANALYSIS REVIEWS 544, 548 (2011) (noting one element of liability is the "expected wrongful harm," where expected is measured in probabilistic terms). I have raised this objection previously. See Kimberly Kessler Ferzan, *Culpable Aggression: The Basis for Moral Liability to Defensive Killing*, 6 OHIO ST. J. CRIM. L. 669, 680 n.44, 681 n.46 (2012) [hereinafter *Culpable Aggression*].

19. See *id.*

20. Kimberly Kessler Ferzan, *Forfeiture and Self-Defense*, in THE ETHICS OF SELF DEFENSE, 233, 252–53 (Christopher Coons & Michael Weber eds., 2016) [hereinafter *Forfeiture and Self-Defense*].

21. *Id.* at 233, 252.

How then should we think about liability cases? First, we should stipulate that liability is simply a form of limited forfeiture or suspension of rights.²² Second, in addition to forfeiture, there need to be positive reasons to impose the harm. This is because we might think that even individuals who are liable to defensive force—and do not have their interests protected by rights—still have interests against being harmed, and one does need a reason to harm those interests.²³ In the standard self-defense case, the very act that causes the forfeiture—a culpable attack—creates the reason to harm the aggressor—to stop that very attack.²⁴ This highlights the benefit of McMahan’s package: because he combines forfeiture with instrumental reasons, his liability formulation should yield that it is permissible to harm people just in those instances in which there is both forfeiture and defensive ends.

The question, then, is what to do when the defender needs to use more force than that to which the aggressor is liable. The answer to this question turns on the boundaries of liability and, specifically, proportionality. Notably, there are substantial differences between my view of self-defense and Steinhoff’s. My hope is to bracket many of these questions. For my purposes in this paper, I think Steinhoff’s position that there is a “proportionality requirement” is all that we need to join issue.²⁵ Although much of his argument for proportionality rests on precautionary principles, Steinhoff is at times sympathetic to a broadly construed “proportionality requirement.”²⁶ “[A] culpable aggressor . . . forfeits rights against proportionate defense . . .”²⁷ Steinhoff believes that a culpable aggressor might be able to complain if his tongue were cut out to prevent an insult.²⁸ To Steinhoff, the severity of the attack is measured by, at the very least, the harm, the culpability,

22. See Steinhoff, *supra* note 1, at 62–63; HELEN FROWE, DEFENSIVE KILLING 72 (2014). Steinhoff is correct that forfeiture is not reason supplying, but he is incorrect to argue that “liability is no path to permissibility at all.” Steinhoff, *supra* note 1, at 38. The fact that no right stands in the way of the action certainly matters to how strong the reasons have to be to justify the action. For further discussion, and amendments to my earlier explication that Steinhoff criticizes, see generally Ferzan, *Forfeiture and Self-Defense*, *supra* note 20.

23. FROWE, *supra* note 22, at 117.

24. See Ferzan, *Forfeiture and Self-Defense*, *supra* note 20, at 244 (arguing that the “facts that ground the forfeiture” are typically the acts that provide the reason to defend).

25. Steinhoff, *supra* note 1, at 103.

26. *Id.*

27. *Id.* at 35.

28. *Id.* at 54.

the threat to the legal order, the threat to autonomy, and the threat to honor.²⁹ In addition, he adopts counterfactual positions—one need not respect rights in others that they would not respect in you were the roles reversed.³⁰ If the apple thief would shoot the farmer dead to prevent the farmer from taking his things, then turnabout is fair play.³¹ And, at times, Steinhoff says proportionality is conventionally constructed.³²

I do not want to defend alternative contours of proportionality here, so let me defend a proportionality requirement on Steinhoff's terms. I am going to assume there is a proportionality requirement that prevents us from killing people or inflicting serious bodily harm for minor infractions. As a somewhat notorious noodge—and sometimes even arm puncher—of my friends, I will simply plead that I would never seriously harm them to prevent them from doing the same to me. So, I do not have the kind of counterfactual disposition that Steinhoff assumes of the apple thief who may be shot by the farmer because he would shoot the farmer were the roles reversed.³³ Sometimes an apple thief is just an apple thief. For purposes of this paper, I am going to assume these actors are also like me and do not have evil counterfactual dispositions. Moreover, given that American jurisdictions tend to place restrictions on the use of deadly force and apply them only to serious crimes, I think there is general agreement that there is a rough and ready proportionality limitation.³⁴

29. *Id.* at 19.

30. *Id.* at 122–23.

31. *Id.* at 122.

32. *See, e.g., id.* at 124, 127.

33. *Id.* at 122.

34. *See* JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW § 18.01[D] (7th ed. 2015) (“[A] person ordinarily is *not* permitted to use *deadly* force to repel what he knows is a *nondeadly* attack.”) (footnote omitted).

This is all the common ground we need. Steinhoff, McMahan, and I are all tempted to use desert-based reasons at times.³⁵ Within the self-defense literature, problems and puzzles have arisen when the forfeiture comes apart from the instrumental reasons. If one is an objectivist, then there is no defensive instrumental good served when defense is futile.³⁶ Even those who take an evidence-relative view of justification face the same puzzles if the defender believes there is no defensive reason for force. Although Steinhoff rejects the former puzzle, he must still confront this one: for those of us who think forfeiture may be limited to only proportionate defense, there is always the problem of what to do when the defensive means available would extend beyond the extent to which the aggressor has forfeited his rights. That is, what may a defender do when the only way to stop an aggressor's pinch is to cut off his arm?

What these questions invite, then, is the inquiry of how we might think about and combine reasons. It comes as no surprise that the fact that an aggressor deserves harm enters the picture.

Let us take a look at punishment. The interaction I am examining is the interaction between *desert* and defense. Hence, we need to look at retributivism.

Punishment, and in particular, retributivism, has features of both forfeiture and reason giving. What is called "negative retributivism" is a forfeiture

35. I have argued that you may not use a person who cannot harm you as a means to prevent harm from another actual threatener. Ferzan, *Culpable Aggression*, *supra* note 18, at 694. This was in response to McMahan's challenge that if all one needs is to be culpable for liability to defensive force, it becomes untethered from a necessity limitation to prevent that attacker's harm. *Id.* I rejected that one could kill such a person as a matter of self-defense, but then suggested it might count as a "lesser evil." *Id.* My frequent co-author has since argued that one may harm such individuals to the extent that they deserve it. Larry Alexander, *Recipe for a Theory of Self-Defense: The Ingredients, and Some Cooking Suggestions*, in *THE ETHICS OF SELF DEFENSE*, *supra* note 20, at 20, 24. I had not considered the punishment angle in my response to McMahan and I concede that Alexander is correct about this. In contrast, Lars Christie has been developing a robust theory of liability wherein he defends what I would not—that one may, as a matter of liability to defensive force, harm attackers to prevent wholly unrelated harms. See Lars Christie, *Harming One to Save Another: Liability and Lethal Luck* (2016) (unpublished Ph.D. dissertation, University of Oslo) (on file with author).

36. For an exploration of this puzzle, see Kimberly Kessler Ferzan, *Defending Honor and Beyond: Revisiting the Connection Between Self-Defense and Success*, 15 J. MORAL PHIL. (forthcoming 2018).

argument.³⁷ Negative retributivism is the view that people can only be punished as much as they deserve; desert is necessary.³⁸ Positive retributivism is a reason-giving argument.³⁹ It is the view—roughly and not uncontroversially—that it is intrinsically good for people to get what they deserve.⁴⁰

The first thing to notice is that forfeiture arguments *constrain/limit* the harm that may be inflicted.⁴¹ Assume A steals a candy bar. Because shoplifting is hard to deter, legal economists conclude that to deter the conduct, A should go to jail for five years. The reason A should not go to jail for five years is because although deterrence is a reason-giving argument, it is not a forfeiture argument. If we assume that desert justifies two weeks of punishment, then A may be punished for two weeks because his negative desert yields that he has no right against being harmed to that extent and it provides a reason to impose that amount of harm.

One initial question is whether, in assessing the reasons *for* punishment, desert shares with other reasons, or whether it occupies the justificatory field.⁴² Consider the objection that punishing Defendant will cause harm to Bystander—perhaps Bystander works for Defendant and will lose her job. A robust retributivist, who takes desert to be the only justifying reason,

37. Michael S. Moore, *Justifying Retributivism*, 27 *ISR. L. REV.* 15, 16 n.2 (1993) (quoting J. L. Mackie, *Retributivism: A Test Case for Ethical Objectivity*, in *PHILOSOPHY OF LAW* 677, 678 (Joel Feinberg & Hyman Gross eds., 4th ed. 1991)). Moore construes negative retributivism as forfeiture; he believes defendants “forfeit such rights precisely by the culpable wrongdoing that constitutes desert.” *Id.* at 36; *see generally* Christopher Heath Wellman, *The Rights Forfeiture Theory of Punishment*, 122 *ETHICS* 371 (2012). Alternatively, Mitchell Berman opts for a specification view. Mitchell N. Berman, *Punishment and Justification*, 118 *ETHICS* 258, 281 (2008) (“The right rearticulation approach I have sketched on behalf of retributivists contrasts with moral forfeiture theory in not suggesting that the offender has forfeited any right. The offender has forfeited nothing.”).

38. Alec Walen, *Retributive Justice*, *STAN. ENCYCLOPEDIA PHIL.* (June 18, 2014), <https://plato.stanford.edu/entries/justice-retributive/>. I will leave to the side the question of whether this view is properly called retributivism or if it is just “Side-Constrained Consequentialism.” R. A. DUFF, *PUNISHMENT, COMMUNICATION, AND COMMUNITY* 11 (2001).

39. *See* J. L. Mackie, *Morality and the Retributive Emotions*, in *WHY PUNISH? HOW MUCH?: A READER ON PUNISHMENT* 271, 272–73 (Michael Tonry ed., 2011).

40. *See id.* With respect to punishment and retributivism, Berman calls the forfeiture-giving argument a condition (Q) and the reason-giving argument (R). Mitchell N. Berman, *Modest Retributivism*, in *LEGAL, MORAL, AND METAPHYSICAL TRUTHS: THE PHILOSOPHY OF MICHAEL S. MOORE* 35, 45 (Kimberly Kessler Ferzan & Stephen J. Morse eds., 2016). He also notes Q is not itself reason-supplying. *Id.*

41. It is of course possible for rights to be overridden. For an alternative to the usual infringing/violating distinction, *see generally* Alec Walen & David Wasserman, *Agents, Impartiality, and the Priority of Claims Over Duties: Diagnosing Why Thomson Still Gets the Trolley Problem Wrong by Appeal to the “Mechanics of Claims.”* 9 *J. MORAL PHIL.* 545, 545 (2012).

42. *See* Berman, *supra* note 40, 43–46 (discussing exclusivity).

must argue that the desert outweighs the harm to third parties. In contrast, as Mitch Berman has argued, a more modest retributivism would share its place on the balance-of-reasons scale with instrumental reasons and allow these reasons to be cashed out against other harms when evaluating whether punishment is supported by the balance of reasons.⁴³ This would mean that the instrumental reasons for punishment can combine and reinforce the desert-based reason.⁴⁴

Assuming desert and instrumental reasons conjoin, it seems they are *concurrently reinforcing*, not *aggregative*. That is, you do not say A deserves X amount of punishment and there are Y instrumental reasons, so A should receive X+Y. Rather, X supports giving the defendant the full amount of punishment he deserves, and instrumental reasons give further weight to that amount.⁴⁵ They do not increase it.⁴⁶

In contrast to the interaction of desert and instrumental reasons, consider the question of how to treat someone who has committed more than one crime. Each criminal act—assuming crimes are correctly defined—gives rise to additional forfeiture and additional positive reasons of desert—to punish. Assume Angela steals five candy bars. Then, it seems that Angela should go to jail for ten weeks, not two.⁴⁷

Though I think this is uncontroversial, let me offer two footnotes to the obviousness of the claim that desert aggregates. First, there are certainly

43. *Id.* at 45–46; *see also* Douglas Husak, *What Do Criminals Deserve?*, in *LEGAL, MORAL, AND METAPHYSICAL TRUTHS: THE PHILOSOPHY OF MICHAEL S. MOORE*, *supra* note 40, at 49, 50–51. Moore agrees with this at the level of institutional design. Michael S. Moore, *Responses and Appreciations*, in *LEGAL, MORAL, AND METAPHYSICAL TRUTHS: THE PHILOSOPHY OF MICHAEL S. MOORE*, *supra* note 40, at 343, 348. But it is not clear to me why we do not need an all-things-considered justification not only for setting up an institution of punishment but also for actually punishing. If terrible things will happen as a result of giving someone what she deserves, that counsels against punishment. And, if there are other consequentialist reasons that are then in favor of punishing in the individual case, I do not see why a retributivist cannot avail herself of these other goods to see what the net result is.

44. This is not to say that instrumental reasons will not cut against giving people what they deserve; just that on the other side of the balance, desert is sharing its spot on the scale with instrumental reasons.

45. For debates about the weight of that reason, *see supra* note 43.

46. This position likely shows my hand that I am not a fan of limiting retributivism and the like. To be honest, I just do not understand the idea that you can deserve a *range* of punishment that can then be determined by instrumental reasons.

47. *See supra* pp. 271–72.

puzzles about volume discounts, so whatever the aggregating function is, it may not be strictly additive.⁴⁸ Second, as a matter of actual legal practice, offenders typically serve the sentences on multiple counts concurrently.⁴⁹ This practice “constitutes a penal sleight of hand.”⁵⁰ Indeed, I would think that the ability to serve multiple sentences at the same time would be an anathema to most retributivists. Desert should aggregate with other desert-based reasons; it should not share simultaneously. *Accordingly, each desert reason is additive as a matter of both forfeiture and reasons.*

We can now ask how desert relates to defense. Consider:

Necessary, but Disproportionate, Defense. A threatens to pinch B on the subway. A is bigger and stronger than B, and B knows the only way to stop A is to tase him. Assume that B’s taser is also such that she knows it delivers a nasty shock but that it does not risk causing A’s death.

Let us add some additional facts. A has previously groped C on the subway. Everyone knows this, including B, but because of a procedural error by the state, A will not be punished for this groping. Let us also assume the amount of punishment deserved by that crime is Y, and Y is equivalent to a punch in the face—which B, being much shorter and weaker will not be able to employ. And, let us assume that tasing A is the amount of force that equals X+Y. May B tase A?

This is just a variation on a hypothetical posed by Jeff McMahan⁵¹ and addressed by Steinhoff.⁵² Here is the question and answer McMahan provides:

Suppose that a person is liable only to X amount of defensive harm but that successful defense requires that X+N amount of harm be inflicted on him. If he

48. That is, we might think that it is two weeks for the first crime, an additional week for the second, four days for the third, and so on. For discussion of volume discounts, see LARRY ALEXANDER & KIMBERLY KESSLER FERZAN WITH STEPHEN J. MORSE, *CRIME AND CULPABILITY: A THEORY OF CRIMINAL LAW* 254–57 (2d prtg. 2011).

49. Julian V. Roberts & Jan W. de Keijser, *Sentencing the Multiple-Conviction Offender: Diminished Culpability for Related Criminal Conduct*, in *SENTENCING MULTIPLE CRIMES* 137, 140 (Jesper Ryberg, Julian V. Roberts & Jan W. de Keijser eds., 2018).

50. *Id.* The only exception to this would be when the delineation of one course of criminal conduct as multiple offenses overstates the defendant’s actual culpability. Then, concurrent sentences undo the problems created by our faulty double jeopardy jurisprudence. For a discussion of the proper delineation of criminal acts, see generally ALEXANDER & FERZAN, *supra* note 48, 226–60.

51. Joshua Glasgow, *Ethics Discussions at PEA Soup: John Gardner and Francois Tanguay-Renaud’s “Desert and Avoidability in Self-Defense” and Jeff McMahan’s “Response,” with Commentary by Victor Tadros*, PEA SOUP (Jan. 10, 2012, 5:26 PM), <http://peasoup.typepad.com/peasoup/2012/01/ethics-discussions-at-pea-soup-john-gardner-and-fran%C3%A7ois-tanguay-renauds-desert-and-avoidability-in-.html> [https://perma.cc/94Q3-G6GZ].

52. Steinhoff, *supra* note 1, at 13.

earlier committed a wrong for which he deserves to suffer N amount of harm, that can make it permissible to inflict X+N amount of harm on him, with the effect of successfully defending his potential victim.⁵³

Steinhoff finds this view to be “odd” based on McMahan’s understanding of liability.⁵⁴ Steinhoff says, “one has to ask where the ‘X’ is coming from. *Ex hypothesi*, X alone has no defensive effect at all.”⁵⁵ There is a way of reading McMahan that does render this conclusion odd. After all, if the defender cannot defend with X, how can the aggressor be liable to it?⁵⁶ Liability, recall, is relative to a goal that can be achieved by the defender’s use of force.⁵⁷ However, there is another way to understand liability, and McMahan’s view thereof, which is fully consistent with finding that there is in fact an X for McMahan. The question parallels one introduced by Jon Quong and Joanna Firth where they imagine Carla is planning to steal Dan’s property and, to do so, she plans to kill Dan.⁵⁸ She is certain to be successful.⁵⁹ Quong and Firth conclude that because any act by Dan is instrumentally ineffective, Carla cannot be liable to harm by Dan.⁶⁰ As Helen Frowe has explained in response:

What Quong and Firth say about this case . . . 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 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. . . . 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.⁶¹

53. Glasgow, *supra* note 51. Robert Nozick previously advanced this argument as well. ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 62–63 (1974) (“One may, in defending oneself, *draw against* the punishment the attacker deserves.”).

54. Steinhoff, *supra* note 1, at 13.

55. *Id.*

56. Admittedly, this is how I saw the problem in an earlier draft. I thank Peter Westen and Doug Husak for pressing me on my too hasty conclusion.

57. *See supra* pp. 268–69.

58. Joanna Mary Firth & Jonathan Quong, *Necessity, Moral Liability, and Defensive Harm*, 31 L. & PHIL. 673, 688 (2012).

59. *Id.*

60. *Id.*

61. Frowe, *supra* note 22, at 97.

Frowe's friendly amendment to McMahan's view would dissolve Steinhoff's problem that there is no X to add. A is liable to X insofar as X is the amount of proportionate harm that would stop the attack. The problem for B is that she lacks the ability to administer X—that is, the punch in the face.

Separate from the interpretation of McMahan, there is the more general question of whether we need this double structure. Steinhoff seems to suggest we do not.⁶² I demur. Even if one does not adopt McMahan's view that necessity is internal to liability—and I am among those who reject this claim—we need an account of why it might be permissible to inflict force that is disproportionate to the harm threatened. So long as Steinhoff grants that there can be a proportionality limitation—that is not grounded in the precautionary principle but rather in the extent of the aggressor's forfeiture—then the harm that it is necessary to impose may outstrip the harm that is permissible.⁶³

May B tase A? There are three possible answers here. First, yes, B is allowed to perform an action that is defensive (X) and it is permissible to give A what he deserves (Y), so X+Y is permissible. Second, no, citizens cannot punish, so B may only employ X. Third, no, like deterrence and desert, X and Y are not additive and only support giving a maximum of Y or X.

I would like us to focus on the first and third answers. Like Steinhoff and McMahan, I believe individuals can inflict punishment in some cases, and thus, I am untroubled by B's doing so in this instance, so I think we can dismiss the second answer for now.⁶⁴ There are, of course, further complications when the state could and would punish, and I will return to those below. But I do not consider desert reasons as unavailable to B.⁶⁵

The first answer is the right one, but let us look at what that entails. Specifically, the forfeiture that supports the desert and the forfeiture that supports the defense are additive—in the same way that multiple criminal offenses are additive. Notice also the odd conjunctions here. The instrumental reason actually supports as much force as is necessary so it outstrips the forfeiture premised on self-defense. At the same time, what does the work here, is the joint forfeitures, plus the desert reason, plus the breadth of the instrumental reason; together they justify imposing X+Y. That is, there is forfeiture for defense (X), forfeiture for desert (Y), a desert reason (Y),

62. Steinhoff, *supra* note 1, at 13.

63. *Id.* at 13–14.

64. *See id.*; Glasgow, *supra* note 51.

65. Admittedly, not everyone shares this view. *See, e.g.,* Malcolm Thorburn, *Judgment, Communication, and Coercion: What's Wrong with Private Prisons?*, 2 *CRITICAL ANALYSIS L.* 234, 243 (2015) (“[I]t is the state, and only the state, that is entitled to interfere with individuals’ basic rights.”).

and an instrumental reason that is equivalent to X+Y as that is the amount of force *necessary* to repel the attack.

This approach better unpacks the reasons at work than one that allows harm only up to X. After all, it seems odd to say that B could impose X, and the state could impose Y, but no one may impose X+Y, as it seems that A cannot complain about ultimately receiving X+Y.⁶⁶

So, if we think that these justifications are additive, what do we think is going on in the typical case? Are both justifications concurrently at work as Steinhoff suggests?⁶⁷ Is the typical case both desert and defense? Consider:

Two for the Price of One? D threatens to punch E on the subway. E can stop D by stomping on his foot very hard with her stiletto heel. Imagine that as she does so, she thinks both that this will serve as effective defense and that D deserves it.

Now, assume a heeled foot stomp is proportionate to a punch (X), and assume a heeled foot stomp is proportionate to D's desert (X).⁶⁸

Recall that, with respect to punishment, instrumental reasons are concurrently reinforcing with desert.⁶⁹ That would mean E is justified in one stomp and that one stomp is both the punishment and the prevention. It is not 2X; it is just X.

66. The only way to say this would be to say that citizens may not punish. Notice that frequently one act gives rise to multiple shifts of Hohfeldian relations, such that we should see it as unproblematic that there is both defense and desert, as opposed to only one forfeiture. For instance, if a man rapes his secretary at work, he could deserve to go to jail, be required to pay her compensation, lose his job, and provide his wife with a good reason to divorce him. There is no reason to think that an action cannot alter several of our rights and duties at the same time and give rise to numerous normative consequences. *See generally* WESLEY NEWCOMB HOHFELD, *FUNDAMENTAL LEGAL CONCEPTIONS* (Walter Wheeler Cook ed., 3d prtg. 1964).

67. *See* Steinhoff, *supra* note 1, at 13–14.

68. Given my own commitments that individuals are not culpable for an attempt until they unleash a substantial and unjustifiable risk, one might wonder whether there is any desert here on my view. Yes, given that D does cause E fear. Although I could tweak some of the hypotheticals either to create last act attempts or to be explicit about the fear causing, I worry that such qualifications to the hypotheticals would be distracting. Hence, I will use hypotheticals that others take to be culpable incomplete attempts, even though I do not, without further qualification. These examples should not be seen as a retraction of my earlier work. *E.g.*, ALEXANDER & FERZAN, *supra* note 48, 197–226.

69. *See supra* text accompanying notes 45–46.

I believe it is implausible that when E acts, even if she acts intending to defend and punish, that her act constitutes an act of prevention and punishment simultaneously. Here is a first stab at the reason why this is so: if a desert reason is not necessary to justify an action, that desert reason is not satisfied and will continue to exist. That is, when an action is overdetermined by the balance of reasons, the desert reason remains inoperative.

Consider a different case. Assume I take my child and a friend to a movie and decide to buy them both ice cream afterwards. The reason that justifies this is something akin to *this is a nice thing to do for both of them*. Now imagine my child also got an A on a recent history test and that I always take him for ice cream to celebrate. I think he would have a complaint against me if I said that the post-movie ice cream trip also counted as his deserved reward for his grade. That is, although I had two reasons on that first occasion to get him an ice cream cone, the fact that one of those reasons was sufficient to justify the conduct led to the other reason continuing to exist. He continues to have a desert-based reason to get a second dessert.⁷⁰

Here is another reason to think that the desert reason remains in force after the self-defensive action is over. If the desert reason, though unnecessary to justify the action, was also satisfied by the action, then we would expect that the state should not punish the aggressor, or punish him as much, when it later has the chance. If stomping on D's foot is both prevention and punishment, then there may be no further punishment for the state to do.

I will call this the idea that there is a "set-off principle."⁷¹ A set-off principle entails that deserved punishment received at t1 can affect what is deserved at t2. There are a number of claims implicit in this set-off principle that ought to be clarified.

70. This case can be constructed to render a different conclusion. Assume an A on a history test is rare for my son *and* that I almost never take kids for ice cream after a movie. I might decide I am going to reward my son *and*—because it would otherwise be rude—to get the other child with ice cream. In that case, because the desert reason is in play, my son would not be entitled to a second dessert.

71. This, too, was suggested by Nozick. NOZICK, *supra* note 53. George Fletcher seems to misunderstand Nozick's claim. He oversimplifies Nozick's contention as arguing "that self-defense should be treated the same as or at least equivalent to punishment [and] that if a defender inflicts harm on an aggressor in order to thwart the aggression, the harm suffered should be deducted from the punishment that the aggressor deserves for his criminal act." George P. Fletcher, *Punishment and Self-Defense*, 8 L. & PHIL. 201, 201 (1989). But Nozick only appears to be considering cases such as *Necessary, but Disproportionate, Defense*, in which one is seeking to harm beyond the proportionate defensive force. *Compare supra* p. 274, *with* NOZICK, *supra* note 53.

The set-off principle is less controversial when one authority's act of punishment impacts the amount of punishment imposed by another authority.⁷² Assume that my child misbehaves in school, and the school calls me to tell me of this behavior and to advise me that they have imposed a punishment because of that behavior. In such a situation, I would ask the school whether the amount of punishment imposed by the school is sufficient to cover the wrongdoing. To be sure, the misbehavior at school is one thing and the defiance of duties to his mother is another. And yet, I suspect that many parents would feel that if the school has punished the behavior, the punishment diminishes—or negates—the amount of punishment the parent ought to impose.⁷³

In the cases under our consideration, the first actor is not an authority.⁷⁴ If anything, the first actor looks like a *vigilante* for imposing the harm. Still, in these cases, we have been assuming that the defender would be acting as a punisher. That is, we are not asking whether the imposition of undeserved harms should be set-off,⁷⁵ but whether prior instances of punishing behavior should be set-off. I find the claim that a defendant has

72. At one point, Steinhoff, who takes a broad view of forfeiture, argues that once someone is outside the moral community, he may be harmed by many people. *See* Steinhoff, *supra* note 1, at 65. I have two responses to Steinhoff. First, one might subscribe to the view that the defendant only forfeits up to his negative desert. Second, the defendant simply does not deserve multiple harms. As Steinhoff is quick to remind us, forfeiture alone justifies nothing. *Id.* at 34. But once the defendant gets what he deserves, desert can no longer justify anything either. So, from Steinhoff's perspective, a defendant who has been punished might not have any right against harm by others, but there would not be any existing *reason* to harm on a desert view. Indeed, Steinhoff seems to reject that individuals who harm a wrongdoer just because they feel like it are doing anything like punishment. *See id.* at 66. But if they are not giving the wrongdoer what he deserves, then there is no reason to harm him.

73. The dual sovereignty doctrine of Double Jeopardy—allowing for federal and state punishment of the same conduct—rejects the position I advocate in the text. Adam Adler has suggested that one can extract a set-off principle from the Due Process Clause. *See generally* Adam J. Adler, *Dual Sovereignty, Due Process, and Duplicative Punishment: A New Solution to an Old Problem*, 124 *YALE L.J.* 448 (2014).

74. *Cf.* Fletcher, *supra* note 71, at 207 (“By like token, those of us who readily reject the ‘down payment’ thesis presumably take the institution of state punishment as the primary form of legitimate violence. . . . Once state punishment becomes institutionalized, philosophers are likely to think that this is the way things must be.”).

75. For a discussion of this wrinkle, see *infra* pp. 286–89.

already been punished by a vigilante and thus deserves less punishment to be entirely plausible.⁷⁶

Indeed, as Doug Husak has persuasively argued, there are reasons to take defendants' claims that they have been "already punished enough" seriously.⁷⁷ Husak nicely deals with a number of arguments that maintain defendants cannot have been punished by public disdain and humiliation following their crimes.⁷⁸ His responses are generally applicable here. First, Husak dispenses with the argument that the punishment is in the wrong mode, as the defendant is not incarcerated.⁷⁹ As he argues, there is nothing about retributivism—or punishment generally—that entails incarceration.⁸⁰ Second, Husak claims the fact that the punishment is not by the state is not disqualifying.⁸¹ Indeed, he notes that given that punishment involves both stigma and hard treatment, the state rarely has direct control over, or gives precise content to, the stigma.⁸² The stigma comes from *the public's perception* of the conviction.⁸³ Hence, for the defendants Husak analyzes, Leona Helmsley and Rabbi Bergman, the stigma that would follow the state's imposition of punishment is precisely the public reaction that precedes the convictions.⁸⁴ Punishment is partially constituted by the community's views of the defendant's treatment, and thus, the state is not the only authority meting out the punishment.⁸⁵ Finally, Husak argues that because hard treatment and stigma are dependent components of punishment, too much stigma by the public should result in less hard treatment by the state.⁸⁶

If the set-off principle is plausible then if instances of self-defensive action are also instances of punishment, the defendant should get a punishment discount at sentencing. That is, if E's stomping on D's foot is both a defensive action and punishment, then D should not be punished as much later. Notably, in cases in which set-off does apply, we might still worry that the prior defensive harming of the defendant cannot fully fulfill the

76. See Douglas N. Husak, *Already Punished Enough*, 18 PHIL. TOPICS 79, 89 (1990).

77. *Id.* at 79, 89.

78. See *id.* at 81, 84, 89.

79. *Id.* at 81–82.

80. *Id.* at 82.

81. *Id.* at 84–85.

82. *Id.* at 85.

83. See *id.* at 86–87.

84. See *id.* at 86.

85. *Id.* at 90.

86. *Id.* at 95–96.

deserved *punishment* because the harm comes without the stigma.⁸⁷ This may place the state in a bit of a Catch-22 because the way that we tend to allow for the expression of stigma is through the harsh treatment. That is, if the defendant deserves stigma but not harsh treatment but the way we stigmatize is typically through harsh treatment, what is the state to do? Hence, the state may lack an effective way to complete the punishment if it cannot impose further harming—in cases in which, for example, a defendant is harmed by the self-defender as much as she would be by the state as punishment.

When an action is fully justified without invoking desert, it seems desert is not in play and later punishment is warranted. This means *Necessary, but Disproportionate, Defense* should come out differently than *Two for the Price of One*? In the former, the desert reason is necessary to justify the use of the taser.⁸⁸ That means A would have a claim not to be punished

87. I thank Alec Walen for this qualification. I do believe, however, that you can give a defendant what he deserves without stigmatizing. Stigma may be constitutive of punishment without being constitutive of desert. For further exploration of this issue, see LARRY ALEXANDER & KIMBERLY KESSLER FERZAN, REFLECTIONS ON CRIME AND CULPABILITY: PROBLEMS AND PUZZLES (forthcoming 2018).

88. Some of the cases imagined by Steinhoff seem to fit within this category. He states that one can never be “*absolutely* certain” one is dealing with a culpable aggressor; therefore, one must be justified in running the risk that one is wrong. Steinhoff, *supra* note 1, at 53. One reason to run that risk is punishment:

If the defender has sufficiently strong reasons to believe that a culpable aggressor would get away without receiving proportionate punishment if he does not inflict this punishment on the aggressor right now, then this might provide a justification to inflict harms on him that are not necessary by the standards of the self-defense justification.

Id. In such a case, were one to believe that moral justification is evidence-relative, then the infliction of the harm is not justified unless the desert reason is employed.

(Y)—or at least subject to hard treatment because Y has already been imposed.⁸⁹

But in *Two for the Price of One*? how do we determine which objective reason is in play if neither operates simultaneously? Let us relax the assumption that the state will not later punish the actor. In the typical case, whatever the defender may think, the state is functioning and the state will later punish the aggressor. Consider then:

Unloaded Gun. F points a gun at G. G knows the gun is unloaded because G took the bullets out. G is not afraid. G punches F in the face. When asked why G did it, G replies, “Because he deserved it for trying to kill me.”

I submit that G was not justified in acting as she did. Why not? Certainly, F forfeited rights against deserved harm and F deserved harm as a result of assaulting G. But we take rule of law values seriously, and G is simply not permitted to *take the law into her own hands* when there is a well-

89. Some theorists justify punishment as a form of self-defense. See, e.g., VICTOR TADROS, *THE ENDS OF HARM: THE MORAL FOUNDATIONS OF THE CRIMINAL LAW* 169 (2011); PHILIP MONTAGUE, *PUNISHMENT AS SOCIETAL DEFENSE* 25 (1995). These views do not ask about the interaction of desert and defense and are thus outside the scope of this discussion. One might think that in such cases there will absolutely be set-off between harms imposed in self-defense and harms imposed as punishment because punishment harms just are self-defensive. That is, at first blush, one might think that if there is an instrumental reason at t1 that justifies force and force is employed at t1, then the reason is used up when it comes time to impose punishment at t2. I think things are more complicated. First, some theorists use theories that are not purely instrumental. For example, Tadros thinks one acquires duties to rescue that, if not fulfilled at t1, require one to do the next best thing—which through some rather elaborate theoretical gymnastics ultimately justify punishment. TADROS, *supra* note 89, at 276. The question under a Tadrosian view will always be about the duty and what a duty to do the next best thing entails. See Kimberly Kessler Ferzan, *Rethinking The Ends of Harm*, 32 L. & PHIL. 177, 192–97 (2012). Say I roll a boulder at Victor thereby acquiring a duty to rescue him. I then try to rescue him, but do not fully succeed—I lose a hand and he loses a leg. It appears that my t2 duty is the unfulfilled duty to rescue his leg, and the fact that I have already lost a hand is largely beside the point. Second, instrumental reasons do not seem to prevent double dipping. What deters aggression? Aggression may be deterred both if the defender fights back at t1 and if the state imposes punishment at t2. Hence, deterring aggression might provide reasons for multiple impositions of harm at different times. For this reason, Adam Kolber may be mistaken in thinking consequentialism can better explain credit for time served than retributivism. See Adam Kolber, *Against Proportional Punishment*, 66 VAND. L. REV. 1141, 1175–76 (2013). Because the benefit of incarceration will be assessed at t1 and t2, there is no reason to assume that there will not be continuing benefit at t2 to imposing further incarceration. Ultimately, because I am a proud card-carrying retributivist, I do not approach punishment this way. But certainly, if one is not asking about the role of desert and its relationship to defense, then one will need to examine how other reasons relate. I hope that what I have said here is thought provoking for such views even if it is not decisive.

functioning state.⁹⁰ The action does not wrong F, but G is ordinarily not permitted to inflict such harm. She is constrained by her social contract.

If this is correct, then in the standard case, *whatever the defender may think*, the only reason that justifies her conduct is the defensive one. This reason—which we are assuming is partially motivating the defender—is operative and sufficient to justify the harm.⁹¹ The state may then inflict deserved punishment another day.⁹²

In contrast, in nonstandard cases, the desert reason may be available. For instance, in *Necessary, But Disproportionate, Defense*, if B were required to abide by the social contract, she would be required to let A wrong her. We might think the wrong to B outweighs her commitment to the social contract, or we might think that the social contract simply releases her from the bar on using force when she must avail herself of the desert reason to defend herself. I have made a similar claim with respect to futile

90. Kimberly Kessler Ferzan, *Provocateurs*, 7 CRIM. L. & PHIL. 597, 600 (2013).

91. John Gardner and François Tanguay-Renaud suggest that there are moral norms of self-defense that exclude reasons of desert, and thus, the reason of desert is defeated:

Probably there are moral norms of self-defense which distinguish the moral position of the self-defender from that of the punisher by excluding reasons of desert from the deliberations of the former and hence, rendering those reasons defeated.

Possibly these moral norms exist for division-of-labor reasons, to keep moral functions apart when their confusion might be a recipe for further injustice. Or maybe the difference runs deeper.

John Gardner & François Tanguay-Renaud, *Desert and Avoidability in Self-Defense*, 122 ETHICS 111, 127 (2011) (footnote omitted). I am skeptical that there is a moral norm of self-defense doing the work here for two reasons. First, it does seem that at times, self-defenders are entitled to invoke norms of desert, as in *Necessary, But Disproportionate, Force*. See *supra* p. 274. Second, and more importantly, I would wonder about the implications of such a norm for *the state*, the subject of the next section. States sometimes have the option to punish or to defend, and it also seems that desert-based reasons are interacting with broad defensive measures in some instances. However, it is possible that my view can be reconciled with theirs with further elucidation on their part. Because this particular claim was not central to their argument, they do not explore it further.

92. I have been presupposing that the defender acts with particular intentions. For me, the relevance of the intention does not relate directly to the permissibility of the action but to the extent of the forfeiture. I do not think culpable aggressors forfeit rights *in rem* but rather only to those who act aware of the need to defend. Accordingly, there is the possibility of mismatches between a mistaken good reason and an actual potential justifying reason. For purposes of this paper, I am simply taking on board that the defender-punisher acts with the relevant intentions. I plan to explore this mismatch puzzle in Kimberly Kessler Ferzan, *Defensive Killing Without Defensive Reason* (manuscript on file with author).

self-defense cases, where the force cannot succeed in averting the injury but may succeed in imposing deserved punishment.⁹³

Let us consider the converse for a moment.

When I Get Out. L is in jail for molesting children. Five months before his release, L begins to plan to kidnap, molest, and kill children. He even draws schematics of how he would outfit a van with cages for them.⁹⁴

Once L has formed the culpable intention to harm children, we might ask whether he is liable to preventive force. Of course he is. But he is already in jail. Assuming the state is preventing L from harming M by keeping L locked up, has L's punishment somehow converted to self-defense? And if so, does that mean L deserves additional jail time tacked on at the end of the prevention?

The state has opted to impose punishment and has publicly announced this. Given that furthering instrumental goals, such as deterrence and incapacitation, were justified—at least initially—by the fact that the defendant had forfeited rights because of his negative desert, this still appears to be a case of punishment.

Notably, early on I said punishment can run concurrently with other instrumental reasons,⁹⁵ so we might say the punishment is also incapacitating L which is stopping him. But here, we think desert plays the primary role in justifying the incarceration.

Of course, if L has served the appropriate time for desert, then the state may still have preventive reasons to intervene against L.⁹⁶ But at the moment, because the state is locking people up and justifying such actions on punishment, not preventive grounds, the desert reason operates first.

Admittedly, there may be some asymmetries here. Citizens ordinarily may not punish, so we take them as defending.⁹⁷ States, on the other hand, can defend and they can punish.⁹⁸ A system of pure prevention detention, I have suggested, would allow the state to punish later. But we may think that *if the state punishes first, it cannot prevent later*. That may be right,

93. See Ferzan, *supra* note 36.

94. Unfortunately modeled after the actual case of Earl Shriener. ROXANNE LIEB, WASH. STATE INST. FOR PUB. POLICY, WASHINGTON'S SEXUALLY VIOLENT PREDATOR LAW: LEGISLATIVE HISTORY AND COMPARISONS WITH OTHER STATES 1 (1996), http://www.wsipp.wa.gov/ReportFile/1244/Wsipp_WashingtonsSexually-Violent-Predator-Law-Legislative-History-andComparisons-With-Other-States_Full-Report.pdf [<https://perma.cc/FMW2-VLEX>].

95. See *supra* pp. 271–74.

96. Moore, *supra* note 37, at 15.

97. See generally Kimberly Kessler Ferzan, *Self-Defense and the State*, 5 OHIO ST. J. CRIM. L. 449 (2008) (comparing the self-defense relationship between citizens and the state).

98. *Id.* at 460.

but we need to be clear about why. Unlike desert which operates retrospectively, defense operates prospectively.⁹⁹ If an aggressor is in jail and while in jail—thereby being prevented as a side-effect of the punishment—the aggressor changes her mind, there would be nothing to prevent later. The defensive reason would disappear. This is not true for *When I Get Out*, in which the liability to defensive force comes from the current intention to do something in the future. In contrast, if the state prevents first, it may still punish later. Although I cannot explore this further here, we might think that the state ought to punish first because the preventive reason might later disappear. Such a position would be a kind of principle of parsimony, wherein we think that states should do as little to harm their citizens as is necessary to achieve its goals.

Thus far I have offered an account that determines which operative reason is by something like institutional role. But, I suspect that institutional role is operating as a proxy for two different ideas. What is at work in the citizen case is that, separate and apart from what the citizen intends to do, citizens ordinarily are not permitted to punish. This means that reason is, *ceteris paribus*, unavailable to the defender. We know citizens defend and do not punish, not because that is what they intend to do, but because citizens have, in the ordinary case, ceded that authority to the state.¹⁰⁰

In contrast, the reason why it matters whether the state is involved in preventive detention or punishment is because we think that punishment ordinarily requires intentional imposition. This requires that we need to know what the harm that the defendant is being subjected to is aimed at doing. Is it a system aimed at averting dangerousness, or is it a system that intends to publicly condemn? If retributive justice is not achieved if the state does not act for desert-based reasons, then we need to know the state's justifying aim to know if justice is being achieved.¹⁰¹

But there is another complicated question in the neighborhood. Assume the defendant could be confined either as punishment or as prevention. What are we to say about a judge who subscribes to side-constrained consequentialism? The judge punishes because there is rights forfeiture—premised on desert—but then for instrumental reasons. After the defendant gets out of jail, could the victim then harm the defendant because he deserves

99. See ALEXANDER & FERZAN, *supra* note 48, at 4.

100. See Ferzan, *supra* note 97, at 471.

101. Cf. Moore, *supra* note 43, at 391, 391 n.178 (comparing “backwards-looking retributive punishment” and “a forward-looking preventative system”).

it? If desert does not share, then does the reason simply float out there hoping to be satisfied?

The victim may not harm the defendant. But we have to figure out why not. One reason why not would be to claim the desert-based reason is satisfied. But I think the other reason would be that—to the extent the defendant forfeited rights against X amount of punishment and the judge imposed X amount—there is no further forfeiture. So, even if there is still a free-floating desert-based reason, it would be rights violating to inflict it.

This answer also points us to a deeper connection between the rationale for the forfeiture and the rationale for imposing the harm. In both the typical self-defense case and the typical punishment case, the reason the aggressor and defendant forfeit rights also provides a reason to inflict the harm.¹⁰² This linkage is not happenstance, and thus, the forfeiture grounding connects with it the operative reason. The reason we do not see this in *Two for the Price of One* is that there are not just two sufficient reasons; there are two sufficient groundings of the forfeiture.

There is one significant counterexample to this, and that is defendants often get credit for time-served for the amount of time they are held in pretrial detention.¹⁰³ Just like the practice of allowing for concurrent sentences, this seems inconsistent with retributivism. As Adam Kolber has noted, this is not intended as punishment and should not count as such, so why the set-off?¹⁰⁴

This issue needs more attention than I can pay it here. Nevertheless, let me say two things. First, I think we are rightly skeptical of whether most pretrial detention is justified.¹⁰⁵ If so, then the question is whether, having unjustifiably detained someone, the state should pay that person back in the same currency of detention.¹⁰⁶ As discussed below, I am somewhat sympathetic to a whole life view of desert, so this does not trouble me. Second, I am just willing to bite the bullet and say *the current practice is wrong!* We ought not to hold people unless we are justified on preventive

102. See Christopher Heath Wellman, *The Rights Forfeiture Theory of Punishment*, 122 ETHICS 371, 373–74 (2012).

103. Kolber, *supra* note 89, at 1175–76.

104. *Id.*

105. See, e.g., RA Duff, *Pre-trial Detention and the Presumption of Innocence*, in PREVENTION AND THE LIMITS OF THE CRIMINAL LAW 115, 121–22 (Andrew Ashworth, Lucia Zedner & Patrick Tomlin eds., 2013).

106. Kolber is certainly correct that we should then pay the acquitted back as well. Kolber, *supra* note 89, at 1151.

grounds in so doing. And when we are justified and engaged in a preventive system, then those days of prevention should not count toward punishment.¹⁰⁷

I would like to introduce just a couple additional wrinkles. Consider first:

Double Slap. N threatens to hit M in the face. M can stop N by slapping him very hard. M slaps N, and the slap stops N's attack. M then slaps N again and says, "this is because you deserve it."

In this case, M imposes harm that cannot be seen as self-defensive. Instead, M punishes N. Is N entitled to set-off? I believe the answer is yes.

As I have noted above, I believe there are good reasons to think that some harms imposed by defenders count as inflicting retributive desert and are thus rightly set-off by the state. There are two questions, though, about viewing *Double Slap* as inflicting punishment that are worth addressing. The first is the question of whether my position is but the first step on the slippery slope to a "whole life view" of desert.¹⁰⁸ I will confess to being more than slightly tempted by such views. If F is in prison for ten years and in that time, heroically saves five people from a prison fire, it strikes me that the currency of desert should allow F to be released early as his reward. Still, I suspect that one could draw a line between those who intend punishment and those who intend harm—punish you versus hurt you—and harm inflicted by individuals and harm inflicted by nature—I hurt you versus a tree falls on you. There is also the possibility of distinguishing between whether someone is justified in imposing a harm that she intends or believes to be punishment and imposing a harm that the attacker believes is unjustified. I take a central worry of this slippery slope to be that we

¹⁰⁷ Kolber states, "I know of no scholar who has argued against it." *Id.* at 1149. Now one has. See also Moore, *supra* note 43, at 391 ("There is no inconsistency in having both a backwards-looking retributive punishment and a forward looking preventative system. Both could be just systems. The only hard question is whether prevention-justified hard treatment, when it has been received first, counts as a credit against retributive-justified hard treatment. . . . If retributivism is the 'jealous mistress' that I have said it is in the past, the answer should be no.") (footnotes omitted).

¹⁰⁸ See Gertrude Ezorsky, *The Ethics of Punishment*, in PHILOSOPHICAL PERSPECTIVES ON PUNISHMENT, at xi, xxiv (Gertrude Ezorsky ed., 2d ed. 1972); see also Larry Alexander, *You Got What You Deserved*, 7 CRIM. L. & PHIL. 309, 317 n.27, 318 (2013). Kolber advances the concern that a set-off principle for preventive detention as then justifying "proportional harsh treatment," which unmoored from the intentionality of punishment leads to many known, but unintended harms, as part of the proportionality calculus. See generally Kolber, *supra* note 89.

quickly move from condoning citizen vigilantes to condoning citizen-criminals harming other criminals. That is, many people shirk at the idea that if I punch Ken, just to hurt him, but he has just that quantum of desert due and owing, that I have acted justifiably.¹⁰⁹

But there are different ways to approach the treatment of the *unknowing punisher*—even if such a person is not a contradiction in terms. First, allow that this is punishment and unproblematic. I think most criminal law theorists would reject that this can count as punishment.¹¹⁰ Second, say that this action is impermissible because even if it is objectively justified as deserved, it is done with the wrong intention. Here, too, many of us will reject the doctrine of double effect to wriggle out of this conundrum.¹¹¹ Third, say Ken has not forfeited rights against the world but only rights against punishment. Although Steinhoff rejects such a view,¹¹² this sort of claim—that we do not have rights only against those who intend to punish us—seems plausible.¹¹³ Hence, we can condemn the unknowing punisher on several grounds. But beyond these moves, there is a fourth move and that is to say that Ken has not forfeited and that I am not justified *but that nevertheless* Ken is entitled to set-off. That is, I acted impermissibly by punching Ken, but he can argue that having been wrongly punched, the harm should be set-off against his future punishments.¹¹⁴ My goal is not to defend one of these views but merely to suggest an array of approaches that can be taken once we grant that citizens can, in some instances, punish. There are various arguments that can prevent such a position from endorsing

109. For further exploration of this puzzle, see ALEXANDER & FERZAN, *supra* note 87.

110. See H. L. A. HART, PUNISHMENT AND RESPONSIBILITY: ESSAYS IN THE PHILOSOPHY OF LAW 4–5 (2d ed. 2008) (defining punishment); Husak, *supra* note 43, at 51 (“A treatment is not punitive because it happens to deprive and condemn. The very *purpose* of a response must be to deprive and to condemn for it to qualify as a punishment. That is, punishments *intentionally* impose condemnation and deprivation.”); Walen, *supra* note 38 (discussing the four elements required for something to be considered punishment).

111. See ALEXANDER & FERZAN, *supra* note 48, at 97–98 (discussing problems with intent-based deontological constraints).

112. Steinhoff, *supra* note 1, at 44.

113. For the debate, see Wellman, *supra* note 102, at 380–84 (surveying the objections to forfeiture approaches that are limited by the punisher’s intention).

114. Doug Husak has suggested to me that the distinction employed by Shelly Kagan between desert reasons and reasons to give someone what they deserve may be a useful way to see this divide. See SHELLY KAGAN, THE GEOMETRY OF DESERT 18 (2012). Imagine L forgets his sleeping child in his car and the child dies. One thing to say is that L has been punished; the other is to say that although L deserves punishment, given his suffering, there are reasons not to give him what he deserves. With respect to the class of cases I am discussing, if one adopts a whole-life view of desert, this distinction disappears.

the view that citizens act rightly when they beat up victims, who unbeknownst to them, are criminals.¹¹⁵

The second question is whether the state truly should employ a set-off principle as it seems to make the state complicit in the wrongful behavior of the victim. It is state-sanctioned vigilantism. I think this objection is wide of the mark. When we consider *Double Slap*, my argument to this point claims both that the first slap was not both defensive and desert-based but was only defensive, and the second slap, as punishment, should be set-off against any future punishment N would receive. The objection, then, is that there is something amiss with this set-off, as if the state is saying, “Thanks for punishing for us.”

But my position hardly entails that claim. First, we might think M acts wrongly by slapping N, even if he does not wrong N. M violates his social contract with the state or undermines rule of law values. This is why most hypothetical cases philosophers employ in which they allow defenders to avail themselves of desert reasons are instances in which they stipulate that state punishment is unavailable.¹¹⁶ That stipulation undercuts the wrong that is done to the state. Second, as Husak notes, the state has at least *prima facie* reasons to use force to stop private stigmatization—and thus punishment.¹¹⁷ If vigilantism is a problem, the state has a reason to stop it. Finally, the fact that the state might recognize past facts should hardly entail that it condones them. And, the opposite position—that the state should turn a deaf ear to those who claim they have already gotten what they deserved—seems to require the state to be oddly indifferent to the appropriate and proportionate response to wrongdoing.

Desert interacts differently with defense than it does with other instrumental goals. This is because defense is not only reason providing but also grounds rights forfeiture. This reveals that in the ordinary case of individual self-defense—far from operating together in the balance of reasons—the sufficiency of defense will render desert inoperative, with the desert reason requiring satiation at another time. Sometimes reasons don’t share.

115. See Steinhoff, *supra* note 1, at 138–39.

116. See, e.g., Ezorsky, *supra* note 108, at xxii.

117. See generally Husak, *supra* note 43.

