

9-14-2018

The Nature of Self-Defense

Samuel C. Rickless

Follow this and additional works at: <https://digital.sandiego.edu/sdlr>

 Part of the [Constitutional Law Commons](#), and the [Criminal Law Commons](#)

Recommended Citation

Samuel C. Rickless, *The Nature of Self-Defense*, 55 SAN DIEGO L. REV. 339 (2018).

Available at: <https://digital.sandiego.edu/sdlr/vol55/iss2/7>

This Article is brought to you for free and open access by the Law School Journals at Digital USD. It has been accepted for inclusion in *San Diego Law Review* by an authorized editor of Digital USD. For more information, please contact digital@sandiego.edu.

The Nature of Self-Defense

SAMUEL C. RICKLESS*

What is self-defense? Most theorists of self-defense are mainly interested in explaining why and when we are morally justified in defending ourselves from a threat posed by another.¹ The moral questions here are important, not just because self-defense represents an interesting moral conundrum, but because morality, at least in this case, is, or should be, a reliable guide to the law. So theorists of self-defense often start with paradigm cases—the culpable aggressor, the justified aggressor, the innocent aggressor, the innocent threat, and so on—and try to explain moral intuitions about them with the help of moral theory, whether Hohfeldian, utilitarian, Thomist, or other.² Progress has been achieved in this way, but, like Uwe Steinhoff, I think it is worth asking the question of what, exactly, is supposed to count as self-defense.³

* © 2018 Samuel C. Rickless. Professor of philosophy, University of California San Diego, and affiliate professor of law, University of San Diego School of Law. Many thanks to Larry Alexander and the USD School of Law for inviting me to the conference that led me to write this Article. Thanks also to Macklin Thornton and the team of editors at the USD Law Review for their diligence and suggestions.

1. See generally, e.g., THE ETHICS OF SELF-DEFENSE (Christian Coons & Michael Weber eds., 2016); HELEN FROWE, DEFENSIVE KILLING (2014); SUZANNE UNIACKE, PERMISSIBLE KILLING: THE SELF-DEFENSE JUSTIFICATION OF HOMICIDE (1st paperback ed., 1996); Jeff McMahan, *Self-Defense and the Problem of the Innocent Attacker*, 104 ETHICS 252 (1994); Judith Jarvis Thomson, *Self-Defense*, 20 PHIL. & PUB. AFF. 283 (1991).

2. See, e.g., Larry Alexander, *Recipe for a Theory of Self-Defense: The Ingredients, and Some Cooking Suggestions*, in THE ETHICS OF SELF-DEFENSE, *supra* note 1, at 20, 25–26; Kimberly Kessler Ferzan, *Forfeiture and Self-Defense*, in THE ETHICS OF SELF-DEFENSE, *supra* note 1, at 233, 248–49; see also Uwe Steinhoff, *Just War Theory: Self-Defense, Necessity, and the Ethics of Armed Conflicts* 19, 160, 164 (2016) (unpublished manuscript) (on file with author) (providing examples of innocent threats, and culpable, justified, and innocent aggressors).

3. Steinhoff, *supra* note 2, at 4–16.

As Steinhoff sees it, the question of the *nature* of self-defense is prior to, in part because it bears on, the question of the *justification* of self-defense.⁴ Understanding the nature of self-defense is a matter of “conceptual analysis of the term ‘self-defense’” that is designed to be “in line with ordinary language use and plausible accounts of self-defense in the legal literature” and that “keeps the investigation of the justificatory conditions of self-defense from being superfluous.”⁵ In these respects, Steinhoff self-consciously follows the lead of Rudolf Carnap, who evaluates conceptual explications in terms of how well they fulfill the following criteria: similarity to ordinary language use, precision, simplicity, and fertility.⁶

Steinhoff’s main thesis is this:

An act token is self-defense if and only if a) it is directed against an ongoing or imminent attack, and b) the actor correctly believes that the act token is an effective form of *resistance* or the act token belongs to an act type that usually functions as a means to resist an attack.⁷

An attack, in this context, is understood, as it is under German law, to be any “threat of violation or actual violation of an interest that is protected by law—that is, of a right—insofar as this threat stems from human action.”⁸ Steinhoff does not *officially* define “resistance,” but some of his comments strongly suggest that he identifies resisting an attack with making it more difficult.⁹ By this, I assume Steinhoff means that what is made more difficult by resistance is the attacker achieving his goal, which may be in the service of another goal, namely the completion of an act that counts as the violation of a right belonging to the victim. Steinhoff also does not officially define what it is for an act to be “directed against” an attack, but he sometimes glosses the concept of *X being directed against Y* as a matter

4. *See id.* at 16.

5. *Id.* at 16, 20.

6. *See* RUDOLF CARNAP, LOGICAL FOUNDATIONS OF PROBABILITY 6–7 (2d ed., 1962). Note, though, that Carnap does not require *exact* similarity—as in the case of explicating the ordinary concept of fish with the technical, biological concept of *piscis*—the former term, but not the latter, applying to whales and seals; and fertility, as he conceives it, requires “useful[ness] for the formulation of many . . . empirical laws in the case of a nonlogical concept.” *Id.* at 5–7. Steinhoff, I surmise, does not really think of conceptual analysis as involving replacement of a Carnapian explicandum by a potentially rather *dissimilar* explicatum that should be useful in *empirical* theorizing: for Carnap, empirical fruitfulness and precision are far more desirable features of conceptual analysis than similarity to ordinary language and moral-theoretical non-superfluity.

7. Steinhoff, *supra* note 2, at 10.

8. *Id.* at 5.

9. For example, Steinhoff writes: “[O]ne has to distinguish between measures that *resist* the attack, make it more difficult, from measures that simply make it more *costly*.” *Id.* at 12. He also says, of a particular example of murder, that when the victim resists, the killer is “forced to increase his efforts to overcome [the victim’s] resistance—this is what *makes* it resistance.” *Id.* at 17.

of *X warding off, or involving the attempt to ward off, Y*.¹⁰ In other places, Steinhoff takes for granted that “fighting back,” “shooting back,” “reflexively punch[ing] [an attacker], as [one has] learned to do in [a] self-defense course,” grabbing an attacker’s wrist and applying all one’s strength to keep his knife away from one’s body, “scratching and biting a rapist,” and sticking a rapist with a needle all count as actions “directed against” an attack.¹¹

Let us begin with Steinhoff’s claim that—if certain other conditions are met—an act directed against an attack is sufficient for self-defense.¹² According to Steinhoff, an attack must be a “threat of violation or actual violation” of a right possessed by the target of the attack.¹³ But rights, as Wesley Newcomb Hohfeld points out, come in different varieties.¹⁴ There are claims, privileges or liberties, powers, and immunities.¹⁵ I am going to assume that Steinhoff thinks that the kinds of rights threatened by attacks are claims, for claims are the kinds of rights that are subject to violation. For X to have a claim against Y that P is for Y to have a duty to X to ensure that P. And Y violates X’s claim against Y that P when Y fails to discharge the relevant duty. So far, so good. But there are different sorts of claims, and claims vary in part according to the kinds of interests they protect. Some claims protect life, others protect bodily integrity, and some protect liberty while others protect property. Is a threat of violation or actual violation of *any sort of claim* sufficient, when the other conditions are met, for an act directed against this threat to count as an act of self-defense on the part of the person whose interest is threatened? I would think not. Certainly, if Jane initiates a threat of violation of my claim to life or bodily integrity or liberty, then I engage in self-defense when I ward off the attack. But if Jane initiates a threat of violation of my claim to *property*, then surely I do not engage in self-defense if I ward off this threat; what I engage in is defense of my property, not defense of my self. So, for Steinhoff’s account of self-defense to match our ordinary language intuitions, the *kind* of attack against which self-defensive action is taken must be restricted to threatened or actual violation of a certain sort of claim, namely

10. *Id.* at 6.

11. *Id.* at 6–12.

12. *Id.* at 76 n.216.

13. *Id.* at 5.

14. WESLEY NEWCOMB HOHFELD, *FUNDAMENTAL LEGAL CONCEPTIONS AS APPLIED IN JUDICIAL REASONING AND OTHER LEGAL ESSAYS* 37 (Walter Wheeler Cook ed., 2d prtg. 1920).

15. *Id.* at 71.

the kind of claim that protects interests that are tied very closely to the person—existence, integrity, and freedom.

Let us now turn to the question of whether the target of an attack counts as engaging in self-defense only if the attack is either ongoing or imminent. It is reasonable to suppose that imminence or ongoingness should play a role in the *justification* of self-defense: there are many reasons to think that self-defense against an attack that is neither ongoing nor imminent is unjustified. But Steinhoff treats the disjunctive *imminent-or-ongoing-attack* requirement as part of the *conceptual content* of the term *self-defense*, and in this, to my mind, he errs.¹⁶

Steinhoff recognizes that some scholars claim that it is possible to stop a future attack in justifiable self-defense long before it is initiated.¹⁷ Paul Robinson, for example, claims that if D knows that X is building a bomb that X plans to detonate at a school, but that it will take X ten days to finish building the bomb, D would be justified in disabling the bomb at any time and need not wait until “the last moment that intervention would still be effective.”¹⁸ In response, Steinhoff claims that this sort of objection is “beside the point”¹⁹:

The question is not whether we could permissibly use force against X, but whether that force should count as *self-defense*. However, it seems to be semantically far-fetched to call the preacher’s drowning Billy the Baby during the baptism—because the Archangel credibly informed the preacher that this is the only way to prevent Billy the Kid from lethally attacking him 20 years later—an act of *self-defense*.²⁰

I beg to differ. If it is a way to prevent Billy the Kid from attacking him in 2036, there is nothing far-fetched about describing the preacher’s drowning Billy the Baby as an instance of self-defense. Indeed, were the case to come to trial, what else could the preacher possibly plead except self-defense? And at trial, I imagine that the prosecutor and the defendant would engage in something like the following exchange:

Prosecutor: Did you drown Billy the Baby?
Preacher: Yes.

16. See *supra* p. 340.

17. Steinhoff, *supra* note 2, at 14.

18. PAUL H. ROBINSON, 2 CRIMINAL LAW DEFENSES 58 (1984). In this case, D engages in *other*-defense, assuming that D does not plan to be at the school at the time X plans to detonate the bomb. But Robinson’s example is relevant to Steinhoff’s analysis of self-defense inasmuch as Steinhoff says he uses “the term self-defense to include other-defense.” Steinhoff, *supra* note 2, at 21. Semantically speaking, of course, this is odd: the term “self-harm”, for example, is not understood to encompass harm to *others*. So Steinhoff’s semantics of *self-defense* is at least partly *stipulative*, a form of analysis that is in some tension with his reliance on ordinary-language intuitions.

19. Steinhoff, *supra* note 2, at 14.

20. *Id.* at 14–15.

Prosecutor: Why?

Preacher: Because I knew that Billy would grow up to attack me twenty years later, and I knew that I would not have the means to defend myself then. So, I figured that the only way to defend myself would be to kill Billy during the baptism.

Similarly, it seems to me semantically acceptable to describe D's disabling of X's bomb ten days before X plans to detonate it as an instance of other-defense, and hence—given Steinhoff's stipulation that self-defense should be understood to include other-defense²¹—as an instance of self-defense. Imminence should be treated as a factor that is relevant to the *justification* of self-defense, but not as a part of the *semantic content* of the word *self-defense*.

Now to a different, and more important, point. Given Steinhoff's definition of an attack as a threat of violation or actual right-violation insofar as it stems from an action,²² it is impossible, conceptually, to engage in self-defense against a threat that does not stem from an *action*.²³ This, I believe, is already problematic. Suppose I am in an elevator with a butcher carrying a large knife who suddenly suffers a series of muscle spasms that could result in the knife being thrust at me with sufficient force to puncture my skin and damage my internal organs. In the sense of *action* that contrasts with something that merely happens to him, the butcher is not *acting*: the muscle spasms are not under the butcher's control any more than the beating of his heart is. Yet if I save myself from serious injury or death by whacking the butcher's arm until he drops the knife, or—if it is the only means to save my life—by killing the butcher, it seems to me that I engage in self-defense. The same would be true if the butcher's primary motor cortex were being electrically stimulated by an evil neuroscientist.

The miscategorization of the butcher case has potentially significant theoretical repercussions. Steinhoff distinguishes both conceptually and morally between “innocent aggressors” and “innocent threats.”²⁴ Innocent aggressors engage in voluntary activity that, if brought to completion, would

21. See *supra* note 18.

22. See *supra* p. 340.

23. Note that Steinhoff suggests that self-defense requires the warding off of an attack that “stems from *human* action.” Steinhoff, *supra* note 2, at 5 (emphasis added). But surely the word *human* here is unnecessary. If Mr. Spock attacks me gratuitously and I try to repel the attack, I engage in self-defense even though the attack was initiated by an alien.

24. *Id.* at 160, 170.

constitute a violation of another's interest. Steinhoff's standard example of an innocent aggressor is "a person who has been drugged against his will and now suffers from paranoid delusions that compel him to kill others because he deems this necessary to save his life from their imagined unjust attack."²⁵ Innocent threats—such as a falling man who poses a threat of harm to a person below him—are, as Steinhoff says, "*passive*": "*Falling from a cliff, being thrown off a cliff . . . are not acts and can therefore not constitute attacks.*"²⁶ The theoretical repercussion of this distinction is that whereas defense against an innocent *aggressor* is technically a form of self-defense, defense against an innocent *threat* is not. The norm that applies to defense against innocent *threats* is not a "self-defense justification," but rather what Steinhoff calls a "justifying emergency justification."²⁷ One major distinction between these two norms is that "there is no self-defense justification for defending oneself against force justified by a self-defense justification, but there is a justifying emergency justification for defending oneself against force justified by a justifying emergency justification."²⁸

Here, then, is why the butcher case matters. According to Steinhoff, a self-defense justification applies only to cases that genuinely fall under the concept of self-defense.²⁹ As Steinhoff regimentes the concept of self-defense, however, resisting the butcher does not count as self-defense, because the motion of the butcher's knife-carrying arm is not an action, or part of an action, and thus the threat to my life does not count as an attack—in Steinhoff's technical sense of *attack*. It follows, according to Steinhoff's theory, that a defensive response to the butcher cannot be governed by a self-defense justification. Now it may well be that the norm governing the butcher case *should* be something like the justifying emergency justification instead. But, importantly, as I have argued, defense against the butcher actually falls under the concept of self-defense. If this is correct, then Steinhoff *should*—to prevent inconsistency—claim that the proper norm for the butcher case is the self-defense justification, which, among other things, does not allow the butcher to legitimately fight back against my attempt to defend myself against him. And yet it is difficult to see why, from Steinhoff's point of view, we should distinguish for moral purposes between the butcher case and the case of the innocent threat. After all, both the butcher and the innocent threat are passive in the relevant sense:

25. *Id.* at 160.

26. *Id.* at 161; *see also* ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 34 (paperback ed. 2013).

27. Steinhoff, *supra* note 2, at 161 (emphasis omitted).

28. *Id.* at 81, 16.

29. *Id.* at 16.

just as the innocent threat cannot help the fact that he is falling, so too the butcher cannot help the fact that he is experiencing muscle spasms.

How should Steinhoff respond to this problem? Well, he might tweak his explication of self-defense by weakening *attack* to *threatening motion of body parts*. Motion of a body part, such as an arm or hand, need not be voluntary in order to be threatening. On this proposal, defense against the butcher *would* count as self-defense. But this proposal is still underinclusive. If you throw yourself down a well at the bottom of which I am trapped, and the result of the fall is that I will be crushed but you will survive, then I surely count as engaging in self-defense if I vaporize you with my ray gun.³⁰ The problem is that you need not be moving any of your body parts: in this case it is your *entire* body that is moving downward, while your body parts might well be immobile. So, we could, on Steinhoff's behalf, replace *attack* with the disjunctive *threatening motion of entire body or body parts*. But this brings us back to the innocent person who was thrown down the well, who did not initiate the motion of his body and who, according to Steinhoff, is not subject to the self-defense norm but rather to the norm of justifying emergency. To solve this problem, then, we must suppose that the threatening motion of an entire body—in order to trigger genuine self-defense on the part of the potential victim—must be *intentional*. Thus, on Steinhoff's behalf, we would need to replace *attack* with *intentional threatening motion of an entire body or—intentional or unintentional—threatening motion of body parts*. But this proposal too is underinclusive. For we can suppose that you are immobile, but that if you continue to stand without moving on the motion sensitive platform where you happen to be located, a bomb next to me will be detonated and I will perish in the ensuing explosion. If you choose to remain immobile and the only way for me to remove you from the platform is to vaporize you, I am surely engaging in self-defense if I do so. The same is true in a situation in which you are keeping your finger immobile on a device that will detonate the bomb near me if you do not move your finger. I suppose Steinhoff could weaken the replacement for *attack* further to *intentional threatening motion or absence of motion of an entire body or intentional threatening absence of motion of body parts or—intentional or unintentional—threatening motion of body parts*. But this, then, raises the question of why there should be one moral norm applying to cases that fall under this doubly disjunctive concept—that is, the norm of self-defense—and another norm—the norm of justifying

30. See NOZICK, *supra* note 26.

emergency—applying to cases that are otherwise relatively similar, namely those involving the *unintentional* threatening motion or absence of motion of an entire body and those involving the *unintentional* threatening absence of motion of body parts. Intuitively it makes sense to suppose that the difference between the two norms—self-defense and justifying emergency—might be grounded in the difference between an attack on the one hand and, on the other, a threat that is not an attack. But, if the concept of self-defense can only be regimented in this doubly disjunctive way, then it seems that we have lost any intuitive way of grounding the difference between the norms.

What is the answer to this conundrum? One possibility we should take very seriously, I think, is that the ordinary language concept of self-defense should not be a guide to moral norms any more than the ordinary language concepts of *doing harm* and *allowing harm*.³¹ The overarching question here is whether an intentional act designed to ward off or mitigate a serious threat to the vital interests of oneself or another is *justified* or *excused*. Some paradigm cases of self-defense involve use of force that is justified, but others, it seems, such as the case of defense against the butcher, involve use of force that is, at best, excused. Whether the use of force to ward off or mitigate a serious threat of harm to oneself or another is justified appears to be determined by whether the person who poses the threat is culpable for doing so. If this is right, then the dividing line for purposes of moral theory should not be tied to the difference between self-defense and justifying emergency, but rather to the difference between culpable and non-culpable threat-imposition.

Let us now move to another serious difficulty. Steinhoff is very concerned to distinguish his account of the nature of self-defense from so-called “instrumentalist” accounts, according to which an act cannot count as self-defense unless it is instrumental “in the sense of being aimed at averting or mitigating an attack or the harm threatened by it.”³² As Steinhoff sees it, neither an act’s being aimed at averting or mitigating an attack, nor its being aimed at averting or mitigating the harm that would be caused by an attack, is necessary for the act’s being an instance of self-defense.³³ Let us consider these claims one at a time.

Does an act of self-defense need to be aimed at averting or mitigating *harm*? Steinhoff says no, and offers, as a counterexample, the case of the old gunfighter:

31. On why these particular ordinary language concepts are only imperfectly connected to moral principles, such as the Doctrine of Doing and Allowing, see PHILIPPA FOOT, *MORAL DILEMMAS AND OTHER TOPICS IN MORAL PHILOSOPHY* 79–80 (2002).

32. Steinhoff, *supra* note 2, at 9.

33. *See id.* at 4–6.

Consider . . . the case of an old gunfighter weary of life. Again and again he is unjustly attacked by young gunslingers who want to build a reputation by killing him. He secretly hopes that one of them will succeed sooner rather than later, but his professional ethics as a gunfighter (or simply habit) does not allow him not to fight back when attacked. So he does fight back, but, given his mood, without the *intention* of averting the harm of being killed. Of course, if he averts the attack, the harm will be averted too, but this is only a side-effect, not an *aim* of his actions. Still, his fighting back clearly counts as self-defense both under law and according to the ordinary use of the term “self-defense.”³⁴

But here we need to distinguish between *harm-types* and *harm-tokens*.³⁵ On the most straightforward interpretation of *Gunfighter*, the old gunslinger is arguably trying to avert this harm-token and that harm-token, while secretly hoping that he will eventually be unsuccessful. Suppose he is attacked by Billy the Kid and intends that Billy succeed in killing him. If that were his intention, then he would be wildly irrational in fighting back. This is because if he aims at a certain end, then, rationally, he should adopt the most efficient means to achieve that end. But fighting back reduces the chances that Billy will kill him, so why fight back—even out of habit or professional ethics—if what he really wants is to commit suicide-by-Billy? So, it is most reasonable to suppose that when the gunslinger fights back he is aiming to avert the harm of being killed *by this particular attacker at that moment*. Consequently, Steinhoff’s example will not work as a counterexample to the claim that self-defense must be aimed at averting or mitigating harm, as long as *harm* is understood as *harm-token* rather than *harm-type*.³⁶

Steinhoff could respond to this objection by insisting that the gunslinger need not even intend to avert the harm-of-being-killed-by-Billy-the-Kid if he fights back:

34. *Id.* at 6.

35. Linda Wetzel, *Types and Tokens*, STAN. ENCYCLOPEDIA PHIL. (Apr. 28, 2016), <https://plato.stanford.edu/entries/types-tokens/>.

36. I am indebted to Dana Kay Nelkin for raising the type-token distinction as a potential source of objection to Steinhoff’s gunslinger example.

Someone who shoots back at an aggressor need not intend to avert a harm any more than someone who puts sugar in his coffee need[s] to intend to avert a bitter taste. You can do both things for all kinds of reasons (for example, because somebody paid for it or because it is customary or a habit of yours).³⁷

I am not convinced that shooting at someone merely because one has been paid to do so or because it is habitual to do so under those sorts of circumstances automatically counts as self-defense. It is true, of course, that an observer watching the gunslinger shoot back would infer that he was doing so in self-defense. But that is really neither here nor there if intentions matter to self-defense, as I believe they do. Certainly, we should be able to agree that if the gunslinger's shooting back was merely an elaborate signal to a confederate to leave town, it would not count as self-defense *even if* it resulted in the incapacitation of the gunslinger's attacker and halted the attack.

But let us suppose, along with Steinhoff, that there are possible cases of self-defense in which the target of an attack does not intend to avert or mitigate *harm*, whether harm-token or harm-type. It might yet be that an act of self-defense must, for conceptual reasons, be aimed at averting or mitigating an *attack*—or something like an attack that does not stem from an action. But this too is something Steinhoff denies, for he claims that “[a]lthough self-defense is directed against attacks it need not be aimed at averting or mitigating them.”³⁸ Surprisingly, Steinhoff uses *Gunfighter* to support this claim as well.³⁹ This is an argument I have trouble following. Again, Steinhoff describes the old gunslinger as hoping that his attacker will be able to dodge the bullets and kill him.⁴⁰ And Steinhoff seems to think this entails that the gunslinger does not intend to avert or mitigate the attack. But Steinhoff also describes the old gunslinger as “trying his best.”⁴¹ So, what exactly, one might ask, is the gunslinger “trying his best” *to do*?

37. Steinhoff, *supra* note 2, at 6–7.

38. *Id.* at 7 (emphasis omitted). Recall that Steinhoff's conceptual-semantic project is to analyze the term *self-defense* “in line with ordinary language use and plausible accounts of self-defense in the legal literature.” *Id.* at 16. However, as Steinhoff recognizes, the claim that self-defense need not aim at averting or mitigating *attacks* diverges from the formulations of self-defense offered by Cicero, Pufendorf, the German Penal Code, and New York State Law, formulations that Steinhoff otherwise treats as paradigmatic. *Id.* at 7–8. Although it may well be, as Steinhoff argues, that his account fits *some* German penal code *commentaries*, one would think that if our analysis of the *concept* of self-defense should be made to fit the legal literature, the fact that Steinhoff's analysis runs counter to most paradigmatic legal reconstructions of the concept should prove dispositive.

39. I say “surprisingly” because Steinhoff describes the section in which *Gunfighter* first appears as arguing for the view “that self-defense need not be aimed at averting or mitigating *harm* but can for example be merely aimed at averting or mitigating the *attack*.” *Id.* at 7.

40. *See supra* p. 347.

41. Steinhoff, *supra* note 2, at 7.

Presumably he is trying his best to aim his weapon at the attacker or at the attacker's weapon. And why is that? Presumably because he is trying to stop the attacker from shooting at him. But that is exactly what averting or mitigating the attack *amounts to!*

Steinhoff goes on to argue that it is possible for a person to engage in self-defense “while *firmly believing* that her actions will not even mitigate the harm or the attack.”⁴² Steinhoff imagines Sally, who has taken a self-defense course, reflexively punching Superman, as she learned to do in the course, while he maliciously pulls her hair, even though she knows that any attempt to defend oneself against a Superman attack is futile. “Intuitively and semantically,” claims Steinhoff, “this is . . . self-defense.”⁴³ But *is* it? Steinhoff recognizes that “not all reflexive or instinctive behavior that on the occasion hinders an attack is defensive.”⁴⁴ For reflexive behavior directed against an attack to be defensive, the relevant reflexes need to have been *trained* or *designed by evolution* for defensive purposes.⁴⁵ In Sally's case, the relevant reflexes were trained to serve defensive purposes, and in the case of a “spider trying to fight off a spider wasp,” the relevant reflexes were designed by evolution to serve the same kind of purpose.⁴⁶

On reflection, however, it seems strange, semantically or conceptually, for an act's status as defensive or not to hinge on the causal history of the dispositions it manifests. Why should this causal history matter? Suppose we discovered that the spider's apparently defensive maneuvers in response to the spider wasp's attack are mere evolutionary byproducts of genetically encoded signaling behavior to other spiders in the presence of spider wasps. Would we simply withdraw our description of the spider as defending itself? I think a number of us would not. This suggests that our descriptive predispositions in these cases are keyed to observable animal *behavior*. And given that we generally take self-defense to involve intentions and beliefs of certain sorts, the most reasonable explanation of our readiness to use the vocabulary of self-defense in describing the spider's activity is that our application of the concept of self-defense is *analogical*. The use of self-defense vocabulary in this case seems no different from the use of words such as *attempt* and *try* to describe the motion of sunflowers or light-sensitive bacteria.

42. *Id.* at 8.

43. *Id.*

44. *Id.* at 14.

45. *Id.*

46. *Id.* at 11.

My suggestion, then, is that our readiness to describe Sally as engaging in self-defense when she punches Superman reflexively is no more than an instance of analogical concept application. To see this, it helps to slow things down. Imagine that, as Sally throws her reflexive punch, time slows down enough for her to become aware of what is happening as her fist is making its way toward Superman's face. In the time she now has to think about what she is doing, Sally recognizes that her reflexive activity is completely useless—indeed counter-productive—for it will do nothing to stop Superman's attack, it may irritate him, and she will likely injure herself in the process. These things being salient to her, I find myself less inclined to describe her reflexive punch as self-defensive. It is, rather, activity that *appears* to be self-defensive without *being* so. The main reason we are tempted to describe her punch as self-defensive is that it is the kind of rapid activity that a person trying to defend herself against attack would or might perform.

Steinhoff appeals to a different example to establish that the futility of action directed against an attack is compatible with the action's counting as an instance of self-defense. Steinhoff calls the protagonist of this next example "Sally."⁴⁷ So, to distinguish this Sally from the Sally in the Superman example, I will call her *Sally2*:

Someone has thrown [Sally2] to the ground and now tries to sink his knife into her chest while she has grabbed his wrist, trying to keep the knife away. For reasons of his own, the attacker is determined to have the knife enter his victim's chest exactly 20 seconds . . . after he has thrown the victim to the ground. He is stronger than [Sally2], so [Sally2] could at best slow down the pace at which the knife is unstoppably approaching her chest. And, indeed, she *would* slow the attacker down if he pushed down the knife with all his strength, but he does not do this because of his 20 second rule. What he does do is adapt his aggressive efforts to her resistance: the harder she resists, the harder he needs to push in order to keep to his schedule.⁴⁸

Steinhoff describes Sally2's action as "futile": she knows that she will not succeed in keeping the knife away, no matter what she does.⁴⁹ Is Sally2 engaging in self-defense? In this case, I believe that the answer is yes. But this affirmative answer, I will now argue, does not support Steinhoff's proposal.

What the Sally2 example shows is that self-defensive action is possible even when it is futile. What the example *does not* show is that self-defense does not require the intention to avert or mitigate an attack, or the harm an attack would cause. Steinhoff describes Sally2 as grabbing the attacker's wrist and "apply[ing] all her strength to keep the knife away" from her body.⁵⁰

47. *Id.*

48. *Id.* at 8–9.

49. *Id.* at 9.

50. *Id.*

Clearly then, Sally2 is trying, and so intending, both to mitigate or prevent the attack *and* to mitigate or prevent the harm that being stabbed would cause—presumably, death. This is front and center in the example. As Steinhoff himself recognizes, not all self-defensive action is successful: sometimes the attacker succeeds.⁵¹ Similar reasoning suggests that if the victim's situation is dire enough, self-defensive maneuvers will almost certainly fail and resistance will be futile.

Steinhoff might reply that the Sally2 example is designed to show, in addition, that self-defensive action is possible even when the target of attack *knows* that it is futile. But, Steinhoff might add, knowledge that a certain result is impossible is incompatible with intending to achieve it: if you know that you cannot jump over the Empire State Building, then it is impossible for you to try or intend to jump over it. You can tense your muscles as if to jump; you can pretend to jump; you can try to jump as high as possible; but you cannot actually try to jump hundreds or thousands of feet in the air from a standing position. So if the target of the attack *knows* that it is impossible for her to stop or mitigate the attack, then she cannot *intend*, or try, to stop or mitigate the attack. And yet it seems possible at the same time for her to act self-defensively. Therefore, self-defense does not *require* an intention to avert or mitigate an attack.

In reply, let me just say two things. First, Sally2 may have excellent reason to believe that her attempt to save herself will fail, because she knows that the likelihood that her assailant will stop or be thwarted is minuscule. But what she does not have is *knowledge* that she will fail to keep the knife away from her body: her assailant may, for all she knows, drop dead of a heart attack before he can press the knife into her flesh, be struck by lightning, or be attacked by a swarm of bees. The example works only if it is realistic; but its very realism is what deprives the victim of the knowledge that her defensive efforts will fail. Second, even supposing that Sally2 *knows* that she cannot possibly stop the assailant from stabbing her to death, she can still try to mitigate or avert the stabbing. What she *cannot* do is *rationally* try to mitigate or avert the stabbing. What this shows is not that intending to stop or mitigate an attack is *impossible* when the victim knows that the intention will not be fulfilled, but that, under these circumstances, self-defense is *irrational*.

Having argued that an act of self-defense need not be aimed at averting or mitigating an attack or the harm it will cause and that a person can engage

51. *See id.*

in self-defense knowing any efforts at averting or mitigating the attack or the harm will be unsuccessful, Steinhoff builds his account of self-defense on the concept of *resistance*.⁵² According to Steinhoff, if an act is directed against an ongoing or imminent attack, then it is sufficient for it to count as self-defensive if *either* the agent correctly believes that the act token is an effective form of *resistance* *or* the act token belongs to an act type that usually functions as a means to resist an attack.⁵³ As we have seen, Steinhoff assumes that resisting an attack is a matter of making the attack more difficult.⁵⁴ Thus, holding an attacker's wrist and employing all one's strength to push away the knife he is holding counts as resistance because the attacker needs to "adapt his aggressive efforts" to achieve his goal, and "an overpowered woman scratching and biting a rapist" is resisting inasmuch as "the rapist is forced to react to her struggle, loos[en]ing his grip around her throat to fend off her scratching hands, keeping his head at a distance in order not to be bitten, etc."⁵⁵

Steinhoff's explication of resistance as a matter of making things more difficult for an attacker threatens to cause problems for his account of self-defense. On the one hand, he tells us, though "Sally's punch against Superman was entirely ineffective, given Superman's superstrength and invulnerability . . . it would still appear to be self-defense."⁵⁶ On the other hand, Sally's punch does not actually make Superman's attack more difficult—in particular, it does not force Superman to adapt his attack to any of her actions. To solve this potential difficulty, and to make sense of the difference between *actual self-defense* and *merely attempting to defend oneself*, Steinhoff inserts a clause that makes it possible for Sally's resistance to count as self-defense: what renders Sally's punch self-defensive is that it is an act "of the right kind," that is, it belongs to an act type—such as punching—that usually functions as a means to resist an attack.⁵⁷

The insertion of this clause has all the markings of ad-hoc-ness: it is designed to solve a very particular problem for Steinhoff's explication of self-defense, and Steinhoff offers no independent principled rationale for it. This should lead us to wonder whether the clause leaves Steinhoff's explication vulnerable to counterexample. Consider then, Sally3, who discovers a strange rock labeled "Kryptonite" in the Joker's lair. When no one is looking, she takes the rock, and when Superman pulls her hair, she takes it out of her pocket and waves it in front of Superman's face.

52. See *supra* note 9.

53. Steinhoff, *supra* note 2, at 8.

54. See *supra* note 9.

55. Steinhoff, *supra* note 2, at 9.

56. *Id.* at 11.

57. *Id.*

Sally3's actions are surely self-defensive, for she knows that Superman's superpowers disappear in the presence of Kryptonite. Unfortunately, the rock is a meteorite, not a piece of Kryptonite, surreptitiously placed in the Joker's lair by Superman himself. The result is that, on Steinhoff's explication, Sally3's actions are not self-defensive. First, Sally3 does not *correctly* believe that waving the rock in front of Superman's face is an effective form of resistance, and second, the act type of rock-waving-in-front-of-a-face does not *usually* function as a means to resist an attack.

Steinhoff claims that his explication of self-defense "solves the problem" of making room for the possibility of *attempted* self-defense: in order to fit our intuitions about self-defense, Steinhoff thinks he must make sense of the possibility of trying unsuccessfully to engage in an act of self-defense.⁵⁸ Steinhoff's reason for thinking that there is a gap between defending oneself and *trying* to defend oneself is that "there is, with *any* action, a difference between trying to do it and doing it."⁵⁹ Steinhoff's explication solves this problem, for there may be occasions when an agent *mistakenly* believes that her act token is an effective form of resistance to an attack *and* her act token does not belong to any act type that usually functions as a means to resist an attack. In these circumstances, if the agent's act is directed against an ongoing or imminent attack, she can be described as trying, but failing, to defend herself against it.

I myself am less sanguine about the possibility of making sense of the distinction between trying and succeeding in the case of self-defense. Notice, first, that it is not *generally* true that there is a difference between trying to do something and doing it. *Mental* actions strike me as the most salient counterexamples to the general claim. For example, I see no daylight between *inferring Q from P* and *trying to infer Q from P*—to try to infer Q from P *is* to succeed in drawing the inference. Similarly, it seems to me that it is impossible to *try* to think of my mother without thinking of my mother. Descartes argues that if I am thinking then I exist.⁶⁰ He *might* have argued that if I am *trying to think*, then I exist. And the reason that inference is valid is because it is also valid to infer that I am thinking from the fact that I am trying to think.

58. *Id.* at 19.

59. *Id.* at 10.

60. RENÉ DESCARTES, RULES FOR THE DIRECTION OF THE MIND: DISCOURSE ON THE METHOD 51 (Elizabeth S. Haldane & G. R. T. Ross trans., 27th prtg. 1984).

The next question, then, is why we should think that self-defense is any different from these mental actions in this respect. Suppose I am doing something that is directed against an attack, in the sense that it is the warding off, or the attempt to ward off, that attack. Suppose further that in doing what I am doing I am trying as best I can to avert or mitigate both the attack and the harm I would suffer if the attack were successful. Under the circumstances, it seems to me that my act is an instance of self-defense: I have a defensive aim, I am doing something to prosecute that aim, and that is good enough. But Steinhoff requires more⁶¹—either some kind of correct belief about whether my act counts as a form of resistance or some kind of relation between my act and some act type that usually functions as a means to resist an attack. As I see it, this is a mistake.

Steinhoff offers us two examples of failed attempts at self-defense. In the first case, Silly Sally intends the act of scratching her nose as a way of defending herself against an armed attack, wrongly believing that the nose-scratching will make the attack more difficult.⁶² In the second case, Silly Ron⁶³ points a wand at an attacker and says *Avada Kedavra* with the intention of killing the attacker with a green light emanating from the wand.⁶⁴ Steinhoff thinks we will agree with him that Silly Sally and Silly Ron are not really engaged in self-defense, although they are trying, unsuccessfully, to do so. But I, for one, do not agree with Steinhoff. I understand these cases as involving strikingly ineffective self-defense, in something like the way in which Sally's punching Superman as a way of stopping him from pulling her hair is strikingly ineffective. In these cases, the agents act with the aim of averting or mitigating an attack. This seems sufficient for their acts to count as self-defensive, even if, unbeknownst to them, their beliefs about the effectiveness of those acts are comically mistaken. This should be no surprise, given the central importance of *intention* to the proper characterization of self-defense: if the nature of self-defensive action consists in doing something with a certain intention, and—as is the case with all or most mental actions—trying to intend entails intending, it becomes impossible to make room between acting self-defensively and trying to do so.

Where, then, does all this leave us? First, I have tried to explain that ordinary language intuitions about the correct application of *self-defense* and its cognates should be treated as a poor guide to moral, and hence, legal, theorizing in this area. This is not because I find Steinhoff's attempt to explicate the concept of self-defense uninteresting or unfruitful. To the

61. See *supra* note 9.

62. Steinhoff, *supra* note 2, at 11.

63. Steinhoff does not use this name, but it seems to fit the case.

64. Steinhoff, *supra* note 2, at 11.

contrary, I find his investigation thought-provoking and helpful, both for purposes of conceptual clarification and for purposes of moral theory. I just think that there are powerful reasons not to tie the self-defense *norm* to the best conceptual clarification of the term *self-defense*. Second, I have argued that there are numerous ways in which Steinhoff's explication of self-defense falls short of its goal, which is to fit both ordinary language intuitions and legal commentaries on the subject. With respect to ordinary language in particular: (1) self-defense concerns threats to claims and, in particular, threats to claims relating to one's person, rather than claims relating to one's property; (2) in the absence of an ongoing attack, imminence of attack is not a requirement for the correct application of the term *self-defense*; (3) as the butcher case shows, self-defense need not be directed against an attack, in the technical sense of a threat of claim-violation—or actual claim-violation—stemming from an *action*; (4) it is actually quite difficult to find acts of self-defense that are not aimed at averting or mitigating harm; (5) reflexive behavior, whether the product of training or evolution, arguably does not qualify as self-defensive, even if it appears to an external observer to be self-defensive; and (6) the proper understanding of self-defense does not leave room for daylight between defending oneself and trying to defend oneself. The proper explication of the concept of self-defense is a matter for further investigation. But I would think that the central core of the concept is simply action that aims at stopping, neutralizing, aborting, averting, diverting, or mitigating a threat to one or more claims relating to one's person—claims to life, liberty, or integrity—that is emanating, or will emanate, from another person.

