

---

JEFF MCMAHAN

SELF-DEFENSE AND CULPABILITY

(Accepted 28 June 2005)

I. THE PROBLEM AND FERZAN'S SOLUTION

Moral agents sometimes have to act on the basis of beliefs that are reasonable in the context but are in fact false. In these circumstances, agents often act in ways that would be right if their beliefs were true but that they would recognize as wrong if they could see that their beliefs were false. Sometimes our tendency is to think that what these agents do is justified—for example, in the case discussed by Ferzan in which one person, Defender, kills another, Threatener, who has loaded one bullet into a revolver, spun the chambers, pointed the gun at Defender's head, and started to squeeze the trigger.<sup>1</sup> We think that Defender was justified in killing even if we discover that the chamber was in fact empty. (Let us refer to this as the “Roulette case.”) We accept, however, that if Apparent Defender had known with certainty that the chamber was empty and that Apparent Threatener would (or could) have pulled the trigger only once, it would have been pointless and therefore wrong to kill her.

In other cases, our inclination is different. In Ferzan's example (call it the “Drama case”) in which one person, Apparent Defender, reasonably but mistakenly believes that

---

<sup>1</sup> I will refer to anyone who faces an initial threat as Defender while anyone who seems to face a threat but is not in fact at risk as Apparent Defender. Similarly, I will refer to anyone who poses an initial threat as Threatener and to anyone who appears to pose a threat but in fact does not as Apparent Threatener. I will also use the terms “Defender” and “Threatener” in cases in which I do not indicate whether the threat is real or merely apparent. In Ferzan's cases Defender is male while Threatener is female. My use of pronouns will track these sex assignments.

he is about to be shot by Apparent Threatener, who is in fact an actor in a play with an unloaded gun, we find it hard to believe that the act of killing is *justified*. How could it be justified—the right thing to do in the circumstances—to kill a morally innocent person who poses no threat to anyone? How can the fact that Apparent Defender has a false belief about Apparent Threatener *justify* action that would clearly be wrong if Apparent Defender had a true belief instead?

Ferzan proposes an ingenious way of discriminating morally between these cases. Following Paul Robinson, she distinguishes between the “triggering conditions” of the right of self-defense and the restrictions on the exercise of that right. She then proposes that we adopt an objective account of the triggering conditions but a subjective account of the restrictions.

On her view, whether it is permissible to engage in self-defense at all depends on objective considerations and is unaffected by Defender’s epistemic condition. But the determination of whether an instance of self-defense is necessary or proportionate has to take into account Defender’s epistemic limitations. Once the triggering condition is met, Defender’s *beliefs* about necessity, probability, and proportionality are authoritative, whether or not they are reasonable in the circumstances. If, for example, self-defense is in the circumstances objectively unnecessary or disproportionate but Defender does not know this, Defender’s self-defensive action may nonetheless be justified so long as the triggering condition has been met.

This view has considerable appeal. Ferzan argues that the triggering condition is Threatener’s culpability. By insisting on the objectivity of this condition—that is, by insisting that Threatener actually *be* culpable rather than permitting self-defense on the basis of Defender’s reasonable but mistaken belief that she is culpable—Ferzan makes it a necessary condition of the justifiability of self-defense that there are grounds for holding Threatener *liable* to self-defensive action.

By contrast, violence directed against those who are not culpable—that is, those who are morally innocent—is not justified, even if it is excused on grounds of nonculpable epistemic error, duress, or compromised capacity for moral agency. This

is true even when the innocent person poses a threat that is both real and wrongful.<sup>2</sup>

What is perhaps most attractive about Ferzan's view is that, by allowing Defender's subjective point of view to have a role in determining whether the restrictions on self-defense are satisfied, it channels any costs of uncertainty or error to the culpable person rather than to the innocent potential victim. The idea is that it is not unjust to Threatener, and does not *wrong* her, if Defender uses unnecessary or disproportionate force in error, provided that Threatener objectively satisfies the triggering condition. For in that case Threatener is, after all, the one who culpably placed Defender in the position in which judgments about necessity and proportionality had to be made.

In short, on Ferzan's account, Defender's epistemic error about Threatener's culpability may defeat a claim of justification, but his epistemic error about necessary or proportionality cannot. Threatener's culpability gives Defender the right to act even in error.

## II. OBJECTIONS TO CULPABILITY AS THE TRIGGERING CONDITION FOR SELF-DEFENSE

### A. *Culpability and the Law*

Despite its appeal, this view does not achieve all that Ferzan wants it to. She wants a view that enables the law to tell people that it is *permissible* to act in self-defense when it is *reasonable* for them to do so. In the Roulette case, for example, Ferzan contends that we do not want the law to tell a person in Apparent Defender's situation that he is justified in acting in

<sup>2</sup> Here Ferzan joins the minority opinion on the permissibility of self-defense against an innocent threatener. Among those who have argued that self-defense is normally impermissible against at least some innocent threats are Noam Zohar, 'Collective War and Individualistic Ethics: Against the Constriction of 'Self-Defense'', *Political Theory* 21 (1993), pp. 606–622; McMahan, Jeff, 'Self-Defense and the Problem of the Innocent Attacker', *Ethics* 104 (1994), pp. 252–290, and *The Ethics of Killing: Problems at the Margins of Life* (New York: Oxford University Press, 2002), pp. 398–411; Otsuka, Michael, 'Killing the Innocent in Self-Defense', *Philosophy and Public Affairs* 23 (1994), pp. 74–94; and Rodin, David, *War and Self-Defense* (Oxford: Clarendon Press, 2002), pp. 79–89.

self-defense but then later condemn his action if his reasonable belief that his life is at risk turns out to be false because the bullet was in fact in another chamber. Ferzan's view yields the intuitively right answer in this case: it implies that Apparent Defender is justified no matter what chamber the bullet is in because Apparent Threatener has acted culpably. There is no reason why Apparent Defender should have to endure any risk as a result of Apparent Threatener's wrongful action.

But now imagine that the situation is exactly the same from Apparent Defender's point of view but that Apparent Threatener is in no way culpable. Suppose that a villain has made it reasonable for Apparent Threatener to believe that she is a participant in a psychology experiment on altruism in which Apparent Defender is also a knowing participant. Her role in the experiment is to load a gun with what she is told is a blank and point it at Apparent Defender's head. Apparent Defender, she has been told, will pretend to be terrified. The aim is supposedly to determine whether a bystander whom they cannot see will intervene. The villain has also made it reasonable for Apparent Defender to believe something entirely different: that Apparent Threatener is a notorious psychopath with a penchant for forcing other people to play Russian roulette. What Apparent Threatener believes to be a blank is in fact a live bullet. But, as in the original Roulette case, it is not in the chamber the hammer would strike.

Can Apparent Defender be justified in killing Apparent Threatener in the original Roulette case but not in this variant, in which everything is exactly the same from his point of view? What should the law say to him? If it is reasonable for him to act in self-defense in the original case, it is equally reasonable for him to do so in this case. We certainly want the law to permit people in Apparent Defender's situation in the original case to act in self-defense. And it cannot say anything different to Apparent Defender in the variant because his situation in that case is, from his point of view, indistinguishable from his situation in the original case. So the law should tell him that it is permissible to kill Apparent Threatener in both. But on Ferzan's account, the triggering condition is not met in the variant, for Apparent Threatener is not culpable. So self-defense is not

justified; therefore the law must initially permit Apparent Defender to kill Apparent Threatener but later condemn his act when it is determined that the bullet was in a different chamber.

### B. *Culpability Without Threat*

Ferzan claims that the appropriate triggering condition for self-defense is culpability. If the culpable agent in fact poses no threat, that is irrelevant to whether Apparent Defender's right of self-defense is triggered or activated; it just means that the necessity restriction is not satisfied. Since on Ferzan's account the restrictions on self-defense are subject to a subjective standard of justification, Apparent Defender can be justified in killing Apparent Threatener if he *believes* that killing is necessary for self-defense.

There are two kinds of case in which self-defense may be unnecessary. In one, there is a genuine threat but there are ways of averting it other than killing or harming Threatener (for example, by incapacitating her or simply retreating from the threat). Ferzan's account is most plausible when applied to this kind of case. If the threat can be evaded without killing Threatener but Defender does not know that, it may seem that his defensive action is justified. One reason for this is that at least some cases of this sort may not really be cases in which self-defense is unnecessary. For there is a clear sense in which self-defense can be necessary even when there is an alternative way of evading the threat, *if* Defender lacks knowledge of the alternative. In the old television show, *Star Trek*, Mr. Spock could subdue and incapacitate an assailant without harming her simply by applying a special Vulcan grip to her shoulder. Suppose that such a tactic has always been possible but we have not known about it. It would be a mistake to say, in light of that, that self-defense against an unarmed assailant has in the past seldom been necessary because, although we did not know about it, the Vulcan grip has always been available to us as an alternative. Certainly it would be a mistake to say that in all those instances in which the Vulcan grip could have been effective, self-defense was unjustified.

The other kind of case, in which self-defense is *clearly* unnecessary, is that in which, though there appears to be a threat, in fact there is none. These cases present greater challenges to Ferzan's account.

Ferzan, following Hurd and Moore (2002, p. 351), contends that risk is an epistemic concept. All risk, on this view, is subjective risk. The objective probability of harm is either 0 or 1. What is true of risks is also true, a fortiori, of threats. In the Roulette case, for example, the bullet is either in the chamber or it is not. Because it is in fact not in the chamber the hammer will strike, there is objectively no risk and therefore no threat. In ordinary language, however, it is natural to say that Apparent Defender does face a threat in this case. So let us change the details so that this becomes a case that is naturally described as one in which there is no threat but only the appearance of a threat. Suppose that Apparent Threatener culpably intends to subject Apparent Defender to a 1 in 6 chance of death but unknowingly grabs a gun that will fire only blanks and inserts a blank mistakenly thinking it is a live bullet. She spins the chambers and puts the gun to Defender's head. The chamber the hammer would strike is empty; it does not even contain the blank.

Ferzan says that the triggering condition is satisfied in this case because Apparent Threatener culpably intends and is indeed attempting to threaten Apparent Defender's life. But if the triggering condition is understood as an *objective* condition for the justification of *self-defense*, it cannot be satisfied in such a case. For this cannot literally be a case of self-defense: there is no threat for Apparent Defender to defend himself against. Although Ferzan writes that in such cases it is Apparent Defender's right to redistribute the risk, there is, by her own account, no objective risk to redistribute.

Let us put this aside. It is a merely semantic point. Ferzan's contention can be reformulated. In the terms I would prefer, her claim is that Apparent Threatener's culpability in the Roulette case makes her *liable* to be harmed by Apparent Defender in what he reasonably believes to be self-defense. Yet in order for the account to have general application, it must answer the question: "Culpability for *what*?" The answer cannot be culpability for an unjust threat, for in the Roulette case

there is no threat. Ferzan's answer in the Roulette case seems to be that the triggering condition is culpability for a wrongful *attempt* (which itself presupposes a wrongful intention) to expose Apparent Defender to a serious risk of death. Suppose, then, that the triggering condition—or, as I would prefer to say, the criterion of liability to defensive force—is culpability for an unjust threat or for an attempted unjust threat to another.

To begin to assess the plausibility of this view, consider what I will call the “Basement Window” case. Defender knows that Apparent Threatener intends to kill him. Because of this, Defender carries a loaded gun. But on one occasion he seizes an opportunity to replace the bullets in Apparent Threatener's gun with blanks. Immediately thereafter in an alley from which there is no ready egress, Apparent Threatener attacks. She fires her gun but because the cartridges are blanks, Defender is unhurt. Puzzled, she continues to fire. At the same time a different and wholly independent sniper, Threatener, appears in a tiny basement window at ground level (it is a dangerous neighborhood). She is about to shoot Defender. Defender has his gun but is not a good enough shot to fire through the tiny window. He can, however, save himself by shooting Apparent Threatener, causing her to drop to the ground in front of the window, thereby blocking Threatener's line of fire.

In this case, Defender knows that Apparent Threatener poses no threat to him. But Apparent Threatener *is* culpable for intending and attempting unjustly to kill Defender. Hence on Ferzan's account the triggering condition has been satisfied. Let us assume that killing Apparent Threatener is the only way that Defender can save himself from Threatener and that because this act would sacrifice one life to save one life, it would not be disproportionate. Given that, on Ferzan's account, culpability for an attempt is a basis for liability when Apparent Defender does not or cannot know that the attempt will fail, the question is whether it can also justify Defender in killing Apparent Threatener as a means of protecting himself from a genuine threat posed by a different person.

The problem is that, vis-à-vis the threat that Defender faces, Apparent Threatener is a mere bystander, indeed an *innocent* bystander—not, of course, *morally* innocent, but innocent in the

sense borrowed from just war theory that makes the term “innocent bystander” redundant. A person who is innocent in this sense is simply a person who poses no threat. Thus, to say that Apparent Threatener satisfies the triggering condition in this case suggests that it can be justifiable for Defender to kill a relevantly innocent bystander *as a means* of averting a threat posed by another. Yet traditional accounts of liability hold that culpability *for a different offense* cannot make a person liable to be harmed as a means of averting a threat for which she is in no way responsible.

The fact that Defender’s killing of Apparent Threatener would not in a literal sense be an act of *self-defense* but would instead be an act of using the killing of Apparent Threatener as a means of *self-preservation* does not seem relevant on Ferzan’s account. For Apparent Defender’s killing Apparent Threatener in the Roulette case is also not literally an act of self-defense, since there is no threat for Apparent Defender to defend himself against. But Ferzan does not take that fact to exclude the possibility of justification.

Ferzan might nevertheless deny that her account implies the permissibility in this case of Defender’s killing Apparent Threatener in self-preservation. For she concedes that “the satisfaction of the triggering condition is not alone sufficient. Self-defense remains limited by [Defender’s] perception of the need to act.” (p. 738) Thus, in discussing a case in which Defender knows that culpable Apparent Threatener’s gun is unloaded and therefore that killing is unnecessary, Ferzan plausibly contends that it is not permissible for Defender to kill Apparent Threatener, even though the triggering condition is satisfied. (p. 738–739 and 746–747).

Yet the Basement Window case is importantly different from the case that Ferzan considers in which Defender’s killing Apparent Threatener would serve no purpose. What is unusual in the Basement Window case is that there *is* a lethal threat to Defender but it does not come from Apparent Threatener but from a different person, Threatener, who poses exactly the *same kind of threat* that Apparent Threatener *attempts* to pose. The difference, in short, is that in Ferzan’s case, killing Apparent Threatener is unnecessary while in the Basement Window case,



Apparent Threatener satisfies the triggering condition and killing her *is* necessary for the defense of Defender's life—it is just that the threat to his life is not posed by Apparent Threatener. In the Basement Window case, both the triggering condition and the restrictions appear to be satisfied.

So it seems that Ferzan's account does, after all, imply that it would be justifiable for Defender to kill Apparent Threatener in the Basement Window case. And perhaps this is not an implication that Ferzan would wish to avoid. For there are various considerations that support the view that Defender would be justified in killing Apparent Threatener in the Basement Window case. There is intuitive support for this view, though intuitions are sufficiently divided here that it would be unwise to rely on them directly. Another line of thought appeals to a parallel with the *punishment* of attempts. One of the acknowledged aims of punishment is deterrence. We punish wrongful attempts in part to deter others from making similar attempts, which might be successful. This means that we accept that those who are culpable for wrongful attempts thereby make themselves liable to harm as a means of protecting the innocent from threats that might be posed *by others*. The idea that it is permissible for Defender to kill Apparent Threatener as a means of defending himself against the threat from Threatener could appeal to the same pattern of justification.

Ferzan might also reflect that if Apparent Defender can be justified in killing Apparent Threatener in the Roulette case, when he actually faces no threat at all, then Defender ought to be justified in killing Apparent Threatener in the Basement Window case. For in the Basement Window case, Defender does face a lethal threat and Apparent Threatener is *more* culpable than Apparent Threatener is in the Roulette case, since in the Basement Window case Apparent Threatener intends only to expose Apparent Defender to a 1 in 6 chance of death whereas in the Basement Window case Apparent Threatener intends to kill Defender. In the Basement Window case, in other words, the triggering condition is more robustly satisfied and the killing of Apparent Threatener would serve an important purpose, namely, saving Defender's life.

### C. *Culpability for What?*

Ferzan's claim that the triggering condition is satisfied in the Roulette case effectively severs culpability from threat. This raises two questions. First, what forms of culpability are sufficient to trigger a right of self-defense (or apparent self-defense)? Clearly not just any form of culpability will do. If I am a pedestrian crossing a road, every driver whose car is moving in my direction poses a threat to my life. The subjective risk is very slight but not negligible. Suppose I could reduce the risk by shooting a particular driver. It would be absurd to suppose that my right of self-defense could be triggered by this driver's culpably lying to his wife over his cell phone, so that the only issue would be one of proportionality. Culpability for lying to his wife cannot make the driver liable to defensive action by me. So, as I noted earlier, in order for Ferzan's account to be determinate in its implications, it cannot claim simply that culpability is the triggering condition: it must also identify the forms and degrees of culpability that are sufficient as triggers and distinguish them from those that are not.

The second question is what kinds of purpose may legitimately be pursued by means of attacking or harming a person who is culpable in the relevant way. I suggested in the previous subsection that if culpability for an attempt justifies killing that is *perceived* as self-defensive but in fact is not, then it might also justify harming or even killing the culpable person as a means of self-preservation from a *different* but *actual* threat that is similar to the attempted threat but for which the culpable person is not responsible.

Yet there is the worry that once the triggering condition is widened beyond culpability for an unjust threat, there may be no principled limit to either the form of culpability or the purpose for which a culpable person may be harmed.

In the previous subsection, I assumed for the sake of argument that the only forms of culpability that trigger a right of self-defense are culpability for an unjust threat and culpability for an attempt to pose an unjust threat. But Ferzan is never explicit in restricting the triggering condition to culpability for threats and attempts. She in fact suggests various extensions of

the triggering condition beyond threats and attempts. She often, for example, equates culpability with the possession of an *intention* to cause wrongful harm. And at another point she suggests that even mere *willingness* to cause wrongful harm could trigger a right of apparent self-defense. Thus, she writes, “anytime that the actor is entitled to make [a] prediction [that harm will occur—that is, any time the triggering condition is satisfied] it will be a fact of the matter than an aggressor is trying to, or at least willing to, injure the defender.” (p. 745) Yet she also notes that “I do not claim that culpability is itself sufficient. Indeed, elsewhere I have specifically endorsed the need for action on the part of the aggressor.” (p. 733) So on her account, mere willingness or an intention to cause wrongful harm is by itself not sufficient to trigger a right of defense—though her own words indicate that she is sometimes tempted to suppose that it might be.

In a note (note 69, p. 737), she indicates that she is also open to the possibility that a wrongful *bluff* is sufficient to trigger a right of apparent self-defense. A bluff involves neither a wrongful attempt at harm nor even a wrongful intention to harm; but Ferzan suggests nonetheless that a culpable act of intentionally creating the misperception of a threat may provide a justification for Apparent Defender to act in apparent self-defense.

What are we to make of this? What wrongful threats, attempts, and bluffs all have in common is that they are intentionally directed at a Defender or an Apparent Defender and may make it reasonable for this person to believe that he faces an unjust threat. If the agent who is responsible for the threat, attempt, or (perhaps) bluff is culpable, she has no ground for complaint if she is harmed by Defender’s or Apparent Defender’s defensive or apparently defensive action.

But if it is culpability and not threat that is the trigger, why should it matter who the intended victim of the culpable action is or even who would be protected by violence directed against the culpable person? To appreciate the force of this question, let us consider three variants of the Basement Window case. These variants involve an expanded cast of characters: Apparent Threatener, Threatener, you, and Victim. The first variant is

just a version of the original Basement Window case involving a third party.

*Basement Window case 2:* Apparent Threatener is culpably attempting to shoot Victim but you have previously loaded her gun with blanks. Threatener is about to shoot innocent Victim from a narrow basement window. You cannot shoot Threatener but can kill Apparent Threatener, thereby saving Victim from Threatener by causing Apparent Threatener to slump to the ground in front of the window.

If Defender is justified in killing Apparent Threatener in the original Basement Window case, you should be justified, for the same reason, in killing Apparent Threatener to save Victim from Threatener. For the difference here is just that between self-defense and other-defense. If the justification for killing in the original case appeals to Apparent Threatener's culpability, there is no reason why that justification should not extend to a third party.

*Basement Window case 3:* Apparent Threatener is attempting to kill you but you have loaded her gun with blanks. Threatener is about to kill Victim by shooting from the narrow basement window. You cannot shoot Threatener but you can save Victim by shooting Apparent Threatener, causing her to fall in front of the window.

If you may shoot Apparent Threatener as a means of saving Victim in Basement Window case 2, it seems that you should be justified in shooting her in this case as well. For what difference could it make to the permissibility of your using the killing of Apparent Threatener as a means of saving Victim from Threatener whether Apparent Threatener's culpable attempt is directed at Victim or at you? In both cases you know that Apparent Threatener poses no threat to anyone.

*Basement Window case 4:* Apparent Threatener is attempting to kill Victim but you have loaded her gun with blanks. Threatener is about to shoot you from the narrow basement window. You cannot shoot Threatener but you can prevent Threatener from killing you by shooting Apparent Threatener as a means of blocking the window.

If you are justified in shooting Apparent Threatener in Basement Window case 3, it seems that you should be justified in shooting her in this case as well. For what difference could it make whose life would be saved from the unjust threat posed by

Threatener, yours or Victim's? If you would be justified in killing Apparent Threatener to save Victim from Threatener, it does not seem that your justification would be affected by Threatener's turning her gun from Victim and training it on you instead.

Now consider carefully the conclusion at which we have arrived: that you are justified in killing a person who you know poses a threat to no one as a means of saving yourself or another from a threat for which the person you would kill is in no way responsible. As I noted earlier, this conflicts with one of the most deeply entrenched principles of common sense morality – namely, that it is wrong to use the serious harming of an innocent person as a means of averting an equivalent or lesser harm to another. As I also noted, it is natural to deny the conflict by claiming that Apparent Threatener is not innocent but culpable. But, again, the traditional prohibition of harming the innocent as a means presupposes a concept of innocence not as *overall moral innocence* but as (i) the state of being harmless or unthreatening (the sense invoked by traditional just war theory) or (ii) the absence of moral responsibility for an unjust threat or other wrong that could be averted or rectified by the act of harming. (In this latter sense, innocence is always relative to a particular act. The same person may be liable relative to one choice but at the same time an innocent bystander relative to a different choice.) Apparent Threatener is innocent in both ways; for she is objectively harmless and is not responsible for any wrong that could be addressed by killing her.

If we say that she may permissibly be killed because she is responsible for a culpable attempt now, we must have a cogent reason for claiming that *that* form of culpability is sufficient for liability (or to trigger a right of self- or other-preservative killing) in these circumstances while other forms of culpability are not – for example, culpability for other forms of wrongful action now (such as lying to one's wife), for a similar attempt in the immediate past, for other forms of serious wrongdoing in the past, for an intention or even mere willingness to engage in an unjust attack against you, or against some other innocent person, and so on. I doubt that any such reason exists. The idea

that the only two triggering conditions are culpability for a threat and culpability for an attempted threat seems arbitrary, or *ad hoc*.

There is, by contrast, a principled rationale for the idea that the criterion of liability to defensive force is moral responsibility for a threat that can be averted only by defensive force. Ferzan, indeed, cites this rationale with approval. It is that if a person has culpably made it the case that either she or someone else must be harmed, then other things being equal (and subject to relevant restrictions such as proportionality) it is permissible as a matter of justice to ensure that she rather than anyone else is harmed. But once we drop the insistence that the person must be culpable *for an actual threat*, and accept that culpability for an attempted threat is also a basis for liability (or triggers a right to harm or kill the culpable person), there seems to be no reason to draw the line at culpability for a present attempted threat. Suppose, for example, that in the original Basement Window case, it is only when Apparent Threatener has thrown down her gun in frustration and is about to walk away that Threatener appears at the basement window and begins to train her sights on Defender. If Defender would have been justified in shooting Apparent Threatener while she was pulling the trigger, surely he would be equally justified in shooting her now, for the same purpose, even though her attempt now lies in the past. Or what if Apparent Threatener had made her attempt yesterday and now just happens to be passing through the alley at the right moment to be used as Defender's shield? Or what if the basis of her culpability is a similar attempt she made a year ago against a wholly different person? And so on.

I do not press this objection merely to try to discredit Ferzan's account of the triggering condition. For I myself am sympathetic to certain extensions of the grounds for liability beyond culpability for a threat. I find it intuitively appealing, for example, to suppose that Apparent Threatener is liable to attack in all the variants of the Basement Window case. If Defender's choice is between killing Apparent Threatener and allowing himself to be killed by Threatener, it does not strike me as wrong to think that he would be justified in sacrificing Apparent Threatener to save himself. It would be another

matter, of course, for Defender to shield himself from Threatener's fire by killing a passerby who was telling outrageous lies to his wife on the cell phone. A principled line must be drawn between Apparent Threatener's culpability and that of the mendacious husband. But I do not know where to draw it.

### III. OBJECTIONS TO THE SUBJECTIVE ACCOUNT OF THE RESTRICTIONS ON SELF-DEFENSE

The cases that Ferzan discusses in arguing for her subjective account of the restrictions on justified self-defense are ones in which a person's culpable action makes it seem that defensive force is necessary when in fact it is not. But necessity is only one of the restrictions on justified self-defense. Another is proportionality. And proportionality in self-defense is not a simple relation between the magnitude of the harm (or expected harm) caused and that of the harm (or expected harm) averted. On any plausible view, what counts as a proportionate defensive response will depend, at least in many cases, on the degree of Threatener's culpability. If culpability is the triggering condition, then because culpability affects proportionality, the restrictions on self-defense cannot be neatly separated from the triggering condition in the way that Ferzan and others assume.

Notice that in Ferzan's examples of justified self-defense, or apparent self-defense, Threatener or Apparent Threatener is highly culpable. This, I think, is what makes it intuitively plausible to suppose that the killing of Apparent Threatener may be justified even when the necessity restriction is objectively not satisfied, as in the Roulette case. But suppose we consider a case in which Apparent Threatener attempts to kill Apparent Defender with a gun that is in fact unloaded but in which her culpability is *slight*—for example, a case in which her reason for trying to kill Apparent Defender is that she believes he is about to commit a murder, though had she been more scrupulous, she could have known that he was an actor rehearsing an outdoor drama. In this kind of case—a minimally culpable attempt—it seems less plausible to suppose that Apparent Defender is *justified* in killing Apparent Threatener rather than merely excused.

Proportionality in self-defense is a function of the relations among three variables: the gravity of the threat Defender faces, the degree to which Threatener is culpable (or, perhaps, responsible) for the threat, and the magnitude of the harm to Threatener that is necessary to avert the threat. On Ferzan's account, Defender is constrained to ensure that his defensive action is proportionate *given* his beliefs about these variables; yet provided the triggering condition is satisfied, his action will be justified even if his beliefs are wrong. (Although Ferzan does not discuss this point, it seems that if the account of the restrictions is fully subjective, it must also imply that Defender's action is *not* justified even if it is objectively necessary and proportionate if he mistakenly *believes* it is unnecessary or disproportionate and acts anyway.) This means that, on Ferzan's account, justification in self-defense is compatible with epistemic error about any of the variables. Or *all* of the variables.

Suppose, for example, that Threatener intends a nonlethal attack. She intends to give Defender a sharp *pinch* in retaliation for an earlier insult. This is unjustified and she has no excuse; she is just in a bad mood. (Call this the case of the "Vicious Pinch.") Defender, however, perceives her assault as posing a *lethal* threat and he therefore perceives her culpability as great. Although he could safely retreat from the confrontation, he fails to notice this. He believes instead that he can save himself only by knocking Threatener unconscious with a blow to the head. But he underestimates the force of his blow, which kills Threatener by crushing her skull.

While Defender believes his defensive action is both necessary and proportionate, he is mistaken about the necessity of defensive action and about the gravity of the threat, the degree of Threatener's culpability, and the magnitude of the harm he inflicts. His action is objectively unnecessary and disproportionate. But on Ferzan's account, because Threatener culpably threatens Defender, the triggering condition is satisfied. This gives Defender the right to act on the basis of his beliefs. Once the triggering condition is satisfied, whether Defender's action is justified depends, on Ferzan's subjective account, not on whether the restrictions are objectively satisfied but on whether



Defender believes they are. Thus, on her account, Defender's killing Threatener is an instance of justified self-defense.

But intuitively this is a clear case of unjustified action. It is true that the burden of choice is thrust on Defender by Threatener's initial culpable action. But her culpability is comparatively trivial. It cannot justify Defender's unnecessary and grotesquely disproportionate action.

Is Defender's action even excused? Intuitively, that depends on whether Defender's beliefs are reasonable in the circumstances. If his beliefs are reasonable, we will see him as excused but not justified. If his beliefs are unreasonable, we will see him as unjustified and culpable. Yet on Ferzan's account Defender's action is justified even if his beliefs are unreasonable. This seems not only intuitively implausible but also unwise from the point of view of the law. For on this view the only belief that Defender has to get right in order to be justified is his belief that Threatener is culpable. Whatever he happens to believe about necessity and proportionality is compatible with justification; therefore he has no moral reason or moral incentive to try to get these beliefs right. But surely he is morally derelict if he makes no effort to test the accuracy of his beliefs; and surely the law ought to encourage him to make the effort.

Although I myself believe that an agent can be liable to self-defensive action without being in any way culpable, I concede that it is not implausible to suppose that culpability for some form of action is the correct criterion for liability to defensive action. As I argued in Section 11, the problem, if one holds this view, is to draw a principled line between those forms of culpable action that are sufficient to trigger a right of defensive action and those that are not. But even if one can solve this problem, what the case of the Vicious Pinch seems to show is that one cannot plausibly hold this objective account of the triggering condition while at the same time endorsing a subjective account of the restrictions on self-defense. It is certainly plausible to regard culpability for certain types of action as sufficient for liability to defensive action, or as sufficient to trigger a right of self-defense. But when the culpability is slight, and especially when the threat is also trivial, it cannot be true that Defender's beliefs about matters relevant to necessity and

proportionality, especially if they are unreasonable, can make his action justified, whatever the facts may be. The threat of a culpable pinch, for example, triggers a right to ward off the pinch by a slap of the hand, or even a slap of the face. But it cannot trigger a right to kill on the basis of unreasonable beliefs.

#### IV. OBJECTIVE JUSTIFICATION

What, then, are we to say about the Roulette case? Is Apparent Defender's killing of Apparent Threatener justified or merely excused? Is Apparent Defender's action in this case really any different morally from his action in the Drama case?

One simple thing to say – the sort of thing philosophers have said before – is that Apparent Defender's action is objectively unjustified but subjectively justified. But suppose we need a univocal concept of justification. On a subjective account of justification – one that allows justification to be affected by Apparent Defender's epistemic situation – Apparent Defender's action is justified. For he reasonably believes that he can avert a 1 in 6 subjective risk of his being unjustly killed by Apparent Threatener only by killing Apparent Threatener.<sup>3</sup> Assuming that justification entails permissibility, subjective accounts imply that Apparent Defender acts permissibly in killing Apparent Threatener. But there is an obvious reason to resist this conclusion. Since Apparent Threatener in fact poses no threat to Apparent Defender, Apparent Threatener has no *reason* to kill her, though of course he understandably believes he does. And it cannot be permissible to kill another person, even a culpable person, for no reason.

---

<sup>3</sup> If the distinction between triggering condition and restrictions can be sustained, subjective accounts can allow the agent's beliefs a role in determining whether the triggering condition is satisfied, whether the restrictions are satisfied, or whether both are satisfied. Subjective accounts can restrict the relevant beliefs to those that are reasonable in the circumstances or, as Ferzan's account of the restrictions does, they can grant a justificatory role even to unreasonable beliefs.

It is tempting to argue that Apparent Defender is *objectively* justified in killing Apparent Threatener in the Roulette case on grounds of just distribution of risk. This approach appears to be an innocuous extension of the account of self-defense to which Ferzan refers according to which justified self- and other-defense is a matter of justice in the distribution of inevitable harm. We simply substitute “risk,” or “expected harm,” for “harm” in our formulation of the account. But at least in the Roulette case, this simply smuggles epistemic considerations in under cover of the notion of risk—as Ferzan rightly observes. (Is there genuine indeterminacy, and hence objective risk, before Apparent Threatener spins the chambers? Even if, as quantum theory suggests, there is genuine indeterminacy in the world, it is unlikely to show up at this level. But if it did, would that mean that Apparent Defender could be objectively justified in killing Apparent Threatener before she spins the chambers but merely excused if he kills her afterwards?)

One alternative is to adhere to a fully objective account of justification. Because in the Roulette case there is in fact no threat to Apparent Defender and thus no reason for him to kill Apparent Threatener, his killing her is objectively unjustified, though fully excused. This of course strikes many people as absurd. But as long as we try to fit Apparent Defender’s act into the standard categories—justified, unjustified and culpable, unjustified but excused—anything we say will seem absurd. Of course it seems absurd to say that Apparent Defender does the wrong thing in killing Apparent Threatener. Do we really believe that someone with a gun to his head and no knowledge of which chamber the bullet is in *ought* simply to do nothing and allow the trigger to be pulled? But it also seems absurd to say that it is justifiable to kill a person who poses no threat to anyone. As Ferzan herself acknowledges, if Apparent Defender *knew* the bullet was in another chamber, it would clearly be wrong for him to kill Apparent Threatener. How, then, can his *ignorance* of morally crucial facts confer justification on an otherwise wrongful act?

I doubt that there is anything difficult or mysterious here at all; it is just that our conceptual resources seem inadequate to

their descriptive task.<sup>4</sup> Here, I think, is what we want to say, though it is difficult to say it in the language of justification. It would be wrong for Apparent Defender to kill Apparent Threatener knowing the bullet was in another chamber. And that wrongful act cannot become right just because Apparent Defender in effect does not know what he is doing. If Apparent Defender kills Apparent Threatener, he has done the wrong thing even though he has acted reasonably in the circumstances. To say that he has acted reasonably is to imply that we would want anyone who has a gun to his head and does not know whether he is about to be unjustly killed to kill the person who apparently threatens him if he can. Morality should tell him, *ex ante*, that he may kill the person who apparently threatens him and the law ought to encourage him to do so.

Yet morality cannot tell him *ex ante* that it is permissible for him to kill Apparent Threatener and then, when it is discovered that the bullet in Apparent Threatener's gun was in another chamber, tell him that he has done the wrong act. He either acts permissibly or he does not. Thus it does not help to distinguish morality's action-guiding function from its deontic classificatory function. These two parts of morality have to speak with a single voice.

If justification is objective—a matter of the way the world is, which is unaffected by erroneous perception or belief—then it may be that when agents act in the face of uncertainty, morality cannot always tell them, at least *ex ante*, what it is permissible to do. It can tell them only what it is *reasonable* to do in their circumstances, given their epistemic limitations. In many cases they may be able to determine, *ex post*, whether their act was justified. If it was not but was nevertheless the reasonable act in the circumstances, they are fully excused and should be evaluated no differently from the way they would have been had their act turned out to have been justified.

As I noted earlier, this strikes most people as a very odd thing to say about the Roulette case. But it is presumably exactly what Ferzan and others who share her intuitions must

---

<sup>4</sup> See Duff, R.A., "Rethinking Justifications," *Tulsa Law Review*, forthcoming.

say about the Drama case. Suppose that from Apparent Defender's perspective in the Drama case the situation is exactly the same as it would be if Apparent Threatener were not an actor but a murderer about to pull the trigger. We want people who are about to be murdered to kill their assailant if that is both possible and necessary for self-defense. Morality tells them that they may do so and the law should encourage them to. But what can morality and the law say to someone who is in fact unthreatened but whose situation is epistemically indistinguishable from that of a person about to be murdered—as is true of Apparent Defender in the Drama case? Morality cannot tell him, *ex ante*, that to kill Apparent Threatener is impermissible. For morality to tell him that, he would have to have knowledge that he cannot have. (Alternatively, perhaps morality *does* tell him that it is impermissible to kill Apparent Threatener, yet this is something that, given his epistemic limitations, he cannot *hear*.) But it also cannot tell him that it is permissible for him to kill Apparent Threatener, for that is false. He may not kill Apparent Threatener in the Drama case because she is innocent and poses no threat to him. Morality can therefore tell him only that it is reasonable for him to kill Apparent Threatener—though it will later have to tell him that what was reasonable turned out to be wrong. If Apparent Threatener were instead a murderer, morality would tell him that it is both reasonable and permissible for him to kill, but unless he could be certain that his assailant was a murderer, the only thing he could *hear* morality say is that killing is reasonable in the circumstances.

I believe that what we should say about the Roulette case is exactly what it seems plausible to Ferzan to say about the Drama case. For what is problematic in both cases is that what it is reasonable for Apparent Defender to do in the absence of relevant knowledge about the apparent threat he faces would be recognizable as wrong in the light of that knowledge. This is a problem about the relation between epistemic limitation and moral justification. This problem is entirely unaffected by the consideration to which Ferzan appeals in order to distinguish the two cases morally—that is, whether Apparent Threatener is

morally culpable (which is also something that normally cannot be known *ex ante*).

Consider next a further problem raised by Ferzan's Variation 4. (p. 721 and 729) Suppose that in the Roulette case a third party knows that the gun that both Apparent Threatener and Apparent Defender believe contains a live bullet in fact contains only a blank. This third party knows that Apparent Defender is not at risk but sees that he is about to kill Apparent Threatener in the belief that this is necessary for self-defense. If Apparent Defender would be, as I have suggested, morally unjustified in killing Apparent Threatener, is it then permissible for the third party to kill Apparent Defender if that is the only way to prevent him from killing Apparent Threatener?

Ferzan rightly judges that this would not be permissible, but the explanation she gives—that the appropriate perspective is that of Apparent Defender, not that of a third party—is not, in my view, the right one. The best explanation appeals directly to the account self-defense as a matter of justice in the distribution of harm. Once Apparent Defender begins to act in perceived self-defense, there is, for the first time, a real threat. It is now true that either Apparent Threatener or Apparent Defender must suffer a harm—in this case, death. For unless the third party kills Apparent Defender, Apparent Defender will kill Apparent Threatener. The question at this point is who among those involved is responsible, or most responsible, for the fact that someone must be harmed. It is true that Apparent Defender is responsible for the threat he now poses to Apparent Threatener, but ultimate responsibility for the situation lies with Apparent Threatener. For the relevant notion of responsibility is *moral*, not merely causal responsibility. And Apparent Threatener not only initiated the crisis but did so culpably, while Apparent Defender is in no way culpable, even though he acts without justification. Thus, given that one of them must be killed, considerations of justice demand that it should be Apparent Threatener.

I will close by posing a challenge to the objective account of justification in self-defense. In the Roulette case, Apparent Threatener inserts a bullet into her gun, spins the chambers,

puts the gun to Apparent Defender's head, and begins to squeeze the trigger. The bullet is not in the chamber the hammer would strike. Suppose that Apparent Defender does nothing and Apparent Threatener pulls the trigger. To which category would the objective account assign this act: justified, unjustified and culpable, or unjustified but excused? Common sense says that the act is unjustified and culpable. But there is no bullet, no threat, no risk. So on what ground is the act unjustified? It is, of course, wrong to frighten people, but in this case that is just a side issue. Presumably it is also wrong to force Apparent Defender to make a momentous life-or-death choice in conditions of great uncertainty. But that too fails to get to the heart of what is objectionable about Apparent Threatener's action.

Some people who accept an objective account of justification are willing to accept that, apart from frightening Apparent Defender and forcing him to make a terrible choice, Apparent Threatener does nothing wrong. If we set aside the lesser wrongs as irrelevant for our purposes, these people would assign Apparent Threatener's act to a fourth category: justified, or perhaps merely permissible, but culpable. They would claim that the act is objectively permissible but reveals something bad about Apparent Threatener as an agent.

I find this hard to accept. It seems to me that Apparent Threatener acts wrongly, that her act is unjustified. One appealing explanation of why her act is wrong is that she acts for a bad reason or with a wrongful intention. The claim that her act is wrong because it involves a wrongful intention is compatible with an objective account of justification. That her act involves a wrongful intention is a fact about the world. Yet many philosophers who embrace an objective account of justification also believe that the intention with which a person acts is irrelevant to the permissibility of her action.<sup>5</sup> The Roulette case seems to me to provide some reason to think that

---

<sup>5</sup> For a brief but forceful defense of this view, see Thomson, Judith Jarvis, "Self-Defense," *Philosophy and Public Affairs* 20 (1991), especially section V. Here she offers a clear example of an act that she believes to be permissible but culpable: a man gives his wife some stuff that he believes to be poison but is in fact the medicine she needs in order to survive.

objectivists about justification would do well to recognize the relevance of intention to justification, and thus to permissibility. But that is an issue for a different occasion.

Department of Philosophy  
Rutgers University  
26 Nichol Avenue  
New Brunswick  
NJ 08901-2882  
U.S.A