04-02-13.2

Recipe for a Theory of Self-Defense: The Ingredients, and Some Cooking Suggestions

by
Larry Alexander

Self-defense and its close relative, defense of others, are uses of force against another person—Attacker—for the purpose of preventing Attacker from harming another's person or property. These uses of force are called self-defense (SD) when employed by the one whose person or property is threatened (the Victim, or V); they are called defense of others (DO) when employed by a non-threatened third party (TP).

SD and DO are preemptive uses of force because they occur before the acts they are intended to prevent occur. They thus belong to the family of preemptive measures that harm people or restrict their liberty in order to prevent them from harming others in the future. Examples of preemptive measures other than SD and DO include preventive detention, restraining orders, gun control laws, and similar laws making possession of certain items illegal for fear of their misuse.

Because SD and DO are preemptive, they operate in the realm of epistemic uncertainty. V and TP can never be certain the feared acts will occur. They can never be certain that if the acts occur, what their consequences will be. They can never be certain what alternative to SD and DO for averting the feared harms exist. And they can never be certain what the status of the presumed Attacker is—that is, whether the presumed Attacker is culpable or nonculpable. I will have had a lot more to say about these epistemic points in due course.

I. The Cast of Characters

A. The Culpable Aggressor

The culpable aggressor, or CA, is one with the following attributes:

(1) The CA intends to commit an act in the future that he believes imposes risks of various magnitudes of harms of various types on V or Vs. And

(2) the CA believes that the facts that will exist at the time he acts, discounted by his estimates of the probabilities of their existence, are such that his imposition of risks of harms will be unjustifiable (regardless of whether he believes it will be unjustifiable).

(3) The CA's culpability is established by (1) and (2). It is not established by the CA's motivating reasons for committing the intended act. Motivating reasons will not render a justifiable act unjustifiable and thus will not render a nonculpable act culpable. However, if the intended act, given the CA's beliefs about risks and circumstances, will be culpable, the CA's motivating reasons can affect the CA's degree of culpability and both enhance it and possibly mitigate it.

(4) The CA's culpability is affected only by his beliefs and not by the evidence for those beliefs. I do not believe negligence is culpable. Therefore, whether or not we think the inferences made by the CA from the evidence he possesses are unreasonable, only the beliefs he forms based on that evidence are material to his culpability.

(5) The culpable act the CA intends may include assisting another CA by aiding him or shielding him.

(6) If the CA will be culpable for acting as he intends, but V or TP know of facts of which the CA is unaware that would make the CA's act justifiable if carried out, then V or TP, as the case may be, will not be justified in using defensive force against CA (although V, but not TP, might be excused).[1] However, the CA will be a culpable person, and that may have implications for how he can justifiably be treated, as we shall see in the next section.

There are two important things to note about CAs as I have defined them. First, CAs can be culpable to varying degrees. The magnitude of their culpability turns on the types and magnitudes of the harms they are risking, on the factual circumstances they believe exist (discounted by the probabilities of those circumstances that they estimate), as well as on their motivating reasons. Motivating reasons may include threats to the CAs that they are attempting to avert but that are not sufficiently serious to excuse their conduct. The magnitude of their culpability may also turn on the quality of their deliberation (for example, whether they are acting impulsively or in the heat of passion).

The second important thing to note about CAs is that their culpability is not the kind of culpability that merits punishment. Kim Ferzan and I in several different works have stressed the distinction between culpable *acts* for which the actor deserves punishment and culpable *intentions* which render the actor liable to have preventive force used against him.[2] We believe

[1] For example, if CA intends to turn the trolley to kill V on the siding, and CA is unaware that there are five trapped workers on the main track who will be saved by his act, then neither V nor his friend TP, who know about the five, will be justified in trying to prevent CA's act. (V may have an excuse or agent-relative permission for doing so, however.) If V does not know about the five, however, then his resistance to CA will be nonculpable, though not, from the perspective of one who knows about the five, justifiable.

[2] *See, e.g.,* Larry Alexander and Kimberly Kessler Ferzan, "Risk and Inchoate Crimes: Retribution or Prevention?," in SEEKING SECURITY: PRE-EMPTING THE COMMISSION OF CRIMINAL HARMS 103 (G.R. Sullivan and I. Dennis, eds., 2012); Larry

someone who intends a culpable act but has not yet committed it is not punishable. We do not believe in inchoate criminality. However, we do believe that by intending a future culpable act, one becomes liable to preventive force.

One final point that follows from the previous one. To be a CA, one does not have to act. One can be entirely passive. Frankie, lying in wait for Johnny, whom she intends to kill, is a CA even if she is not moving a muscle. More on this in Section III.

B. The Culpable Person

The culpable person or CP is someone who has committed a culpable (in the sense of punishable) act or acts in the past. The CP may have acted culpably towards V in the immediate past (for example, he shot at V but missed, or wounded but did not kill V). Or the CP may have acted culpably towards others but has not been punished yet. (One who has been punished for his culpable acts is no longer a CP.) The CP may be one who culpably gave aid to the CA. (If he is culpably intending to give aid to the CA but has not yet done so, he is himself a CA, not a CP.) Or the CP may have had nothing to do with the CA or the CA's attack.

A key question regarding CPs is whether they can be used as means by TP or V to prevent CA's attack and thereby harmed up to the amount they deserve by virtue of their unpunished culpable acts. Suppose CP has tried to kill V and failed, and CP is now out of ammunition. May V grab CP as a shield against CA's murderous attack? May TP shove CP into the path of CA's spear?[3]

Alexander and Kimberly Kessler Ferzan, "Danger: The Ethics of Preemptive Action," 9 OHIO STATE J. OF CRIM. L. 637, (2012).

[3] Although others may wish to distinguish CPs who are causally responsible for V's plight and CPs who are not, I would not do so. In neither case is V defending against the CP. It is CP's negative desert that is the basis of his liability to be used as a means of defense, not his involvement with the particular V who is using him.

Kim Ferzan rejects the proposition that CPs may be used as means. But if their use as means would subject them to no more harm than they deserve for their unpunished culpable acts, would it not be just to so use them? Does the deontological constraint against using people—their bodies, labor, and talents—as means to others' welfare apply to the culpable? Victor Tadros believes some CPs may be used as means for purposes of deterrence because they have acquired a duty to allow themselves to be so used.[4] I believe that all CPs may be used as means to prevent others' harms if in so doing they suffer no more than they deserve.[5]

A second question is whether CP's interests may be discounted in a lesser evils balance. If two villains murderously attack V, but one is now out of ammunition and is thus a CP, may V or TP, in using force against the remaining CA, discount CP's interest in not being harmed as collateral damage and give it less weight than they would give the interests of innocent bystanders? In such a case the CP would not be being used as a means. And intuitively it seems plausible that the TP and V, in defending against a CA, may permissibly use less care to avoid harming a bystander if the bystander is a CP rather than an innocent. Again, I believe this is because a CP deserves to suffer.

C. The Culpable Faker

The culpable faker (CF) is one who (1) is aware that he is creating a risk of a certain magnitude that he will be viewed by V to be a CA, and (2) is aware of no facts, discounted by his

[4] *See* VICTOR TADROS, THE ENDS OF HARM (2011).

[5] On this point, I am in apparent disagreement with my frequent co-author, Kim Ferzan. *See* Kimberly Kessler Ferzan, "Culpable Aggression: The Basis for Moral Liability to Defense Killing," 9 OHIO ST. J. CRIM. L. 669, 694 (2012). Ferzan would, however, hold out the possibility that V's use of a CP as a means for defending himself might be excusable (or agent-relative permitted) even if not agent-neutrally justifiable. Jeff McMahan, on the other hand, seems more sympathetic to my view that the CP's unpunished culpability renders the CP liable to be justifiably used as a means. *See* Jeff McMahan, "Individual Liability in War: A Response to Fabre, Leveringhaus, and Tadros," 24 UTILITAS 278, 285–87 (2012).

estimate of their probability, that would justify his creating that risk. A person who points what he knows to be an unloaded gun at V and threatens to kill him is a CF if he is aware of no facts that would justify scaring V this way.

A CF is liable to be treated as if he were a CA unless and until he is discovered to be a CF. From TP's perspective, moreover, if TP realizes that CF is a CF, but V does not, V should be regarded as an innocent aggressor vis-à-vis CF. As I shall argue below, innocent aggressors should be treated on a lesser evils model by TP, which in the case of CFs should mean that the interests of V as an innocent aggressor should prevail over those of culpable actors, including CF. If TP has no choice other than employ deadly force to prevent V from using deadly force against CF, TP must refrain from doing so.

More controversially, however, when V threatens preventive force against the CF, and the CF cannot credibly communicate to V that he is a CF and not a CA, the CF is now in the position of facing an innocent aggressor (V). As I shall argue below, V's using preventive force against innocent aggressors may be deemed excused on a duress model (or, alternatively, be deemed an agent-relative permission) for the use of that force. Thus, although CF should be punishable for the culpable fakery—and punished severely if he was aware that he risked creating a deadly dilemma—he should not be punished for his defense against V if it was unavoidable and V's threat would otherwise have been sufficient for a duress defense for CF.[6]

D. The Anticipated Culpable Aggressor

[6] *See also* Ferzan, *supra* note 5, at 680 n. 44. If the CF has culpably risked the deadly dilemma of his life or V's (where V, believing CF to be a CA, is prepared to use deadly force against CF), then even if CF is excused (or agent-permitted) to use deadly force against V (if CF cannot retreat or credibly communicate that he is faking), CF will deserve a severe punishment. Indeed, because I do not believe the results of culpable acts should matter for punishment, CF will deserve that same severe punishment even if the deadly dilemma is averted.

The anticipated culpable aggressor (ACA) is one who is predicted (by V or TP) to be a CA in the future. An example: V has had an affair with ACA's wife, and ACA has often said that he will kill anyone who cuckolds him. V knows that ACA's wife has left a voicemail message on ACA's office phone confessing the affair, and he knows that ACA keeps a gun in his office. ACA is about to enter his office and play his voicemail. V, who is in a wheelchair because of knee surgery, sees ACA about to enter his office. V predicts ACA will be a CA when he comes out of his office. V's best chance to defend himself is before ACA enters, not after he emerges as a CA. (I shall discuss the liability of ACAs when I discuss anticipated innocent aggressors.)

E. The Innocent Aggressor

The innocent aggressor, or IA, is one who, from the vantage point of either V or TP or both, is about to impose risks of harms of various types and magnitudes on V nonculpably, risks for which, based on V's or TP's view of the facts (their beliefs about the facts discounted by the probabilities they assign to those beliefs), there is no justification for imposing. An IA may be nonculpable for any of several reasons. The IA may be nonculpable because he may believe the risks he is imposing are less substantial than the V or TP believes they are. (It is immaterial what evidence the IA possesses or whether one might deem him "negligent" for holding his beliefs given his evidence; for as I said in discussing the culpability of the CA, I do not regard negligence to be culpable.) Or the IA may be nonculpable because he may believe, contrary to what V or TP believe, that facts exist that would render imposing the risks justifiable. (Again, his evidence for his beliefs is immaterial.) Or the IA may believe, as do V and TP, that he is about to impose unjustifiable risks on V, but, he is acting under duress and would be nonculpable for that reason. Finally, the IA may be nonculpable because he is not a responsible agent (due to insanity, infancy, or altered consciousness caused by hypnotism, somnambulism, automatism, or the like, or because he is not an agent at all but is, for example, a mere projectile being hurled at V).[7]

Because IAs are not culpable, they are not liable to the use of preventive force. Use of preventive force against IAs must either be justified on a lesser evils model or else be excused (or agent-relative permitted) on the same basis as acts committed under duress. The TP may use preventive force against an IA only if doing so is the lesser evil, as the TP will not be acting under threat to himself or his family. V's use of preventive force against an IA may be either permissible on the lesser evils model or impermissible but excusable under the model of duress. (I take no position here on whether duress is best conceptualized as an excuse or as an agent-relative permission. Either way, it can be invoked only by V, not by TP.)

1. McMahan on IAs

Jeff McMahan dissents from my views on IAs. He believes some IAs *are* liable to preventive force, and not only when and because such use fits the lesser evils paradigm. For he believes some IAs are liable because they are causally responsible for the predicament faced by V and his friend, TP. The example that motivates McMahan's view is one in which a careful (nonculpable) driver—the IA—loses control of his car, which is now bearing down on pedestrian V. V has a weapon that will destroy the car and avert its hitting him but will also kill the IA. McMahan argues that the IA is liable to being killed thusly and that, therefore, V's killing IA would be permissible. The reason McMahan gives is that the IA, although not culpable, is

[7] I am therefore treating the so-called Innocent Threat as just another IA.

morally responsible for the predicament because the predicament was a foreseeable consequence of his choice to drive. (Nonculpable accidents happen.)[8]

The problem with McMahan's view is that it overlooks the reciprocal, Coasean nature of the risks imposed in this case and in the other examples McMahan employs. If V, the pedestrian, is carrying a device which is capable of blowing up cars that veer off the road at him and is prepared to use it, his choosing to walk near the road imposes a foreseeable risk on innocent drivers like IA. (As Kim Ferzan points out in making the same argument against McMahan, imagine that V plants mines alongside the road he walks near in order to blow up any car that goes out of control and would otherwise pose a threat to him. Who, then would be imposing foreseeable risks on whom?)

The flaw in McMahan's position on IAs can be seen as well in his examples of IAs who in his opinion are *not* liable to preventive force. Representative of these IAs is someone who is about to operate his cell phone, which, unbeknownst to him, has been rigged to set off a bomb that will kill V. McMahan argues that this IA is not liable to preventive force because choosing to use a cell phone is, unlike driving, not a choice to engage in a risky endeavor.[9] The problem, however, is that this is not true. Using *this* cell phone *is* a risky choice. IA does not realize how risky it is, but neither does the conscientious driver IA. Moreover, if McMahan can imagine a cell phone rigged to detonate a bomb, so can any cell phone user. Ordinarily, drivers will probably estimate the risk of losing control as greater than the risk most cell phone users would estimate for their cell phones being rigged to bombs—though if one is an undercover agent in a terrorist cell, one might estimate the latter risk to be quite substantial. Yet, it is stipulated that

[8] *See* Jeff McMahan, "The Basis of Moral Liability to Defensive Killing," 15 PHIL. ISSUES 386, 393–98 (2005).
[9] Id. at 398.

the risk from the driver's perspective is low enough to deem him nonculpable and thus an IA and in that sense indistinguishable from the cell phone operator. (The actual "risk" in both hypotheticals is, of course, 100 percent, given that the driver *does* lose control and the cell phone *is* rigged to a bomb; but the believed risk, not the actual risk [of either zero or 100 percent], is what affects culpability.) The cases of the conscientious driver and the cell phone operator are perfectly symmetrical. McMahan's argument for making some IAs liable to preventive force fails.

2. Quong on IAs

Jonathan Quong also rejects McMahan's account for why some IAs are liable to preventive force. His objections to McMahan's account are similar to Ferzan's and mine. But he also denies that only CAs are liable to preventive force. Instead, he argues that some IAs are liable because they are behaving as if others are liable to the harms they risk imposing or as if their victims lack the stringent moral claims that people normally possess. Quong calls his account of liability to preventive force the "moral status account."[10]

Quong rejects restricting liability to CAs for two reasons. First, he thinks that if there are degrees of culpability, which there surely are, and if the degree of preventive force to which one is liable is a function of one's degree of culpability, counterintuitive results follow.[11] I shall take up this argument of Quong's when I discuss the requirement that preventive measures be proportional.

[10] Jonathan Quong, "Liability to Defensive Harm," 40 PHIL. & PUB. AFF. 45 (2012).

[11] Id. at 50–53.

Quong's second reason for rejecting restricting liability to preventive force to CAs is that he believes so restricting it produces the wrong verdict in the following kind of case, which he calls *Mistaken Attacker*:

> The identical twin brother of a notorious mass murderer is driving during a stormy night in a remote area when his car breaks down. Unaware that his brother has recently escaped from prison and is known to be hiding in this same area, he knocks on the door of the nearest house, seeking to phone for help. On opening the door, Resident justifiably believes the harmless twin is the murderer and lunges at him with a knife (Resident has been warned by the authorities that the mass murderer will almost certainly attack anyone he meets on sight.)[12]

Quong argues that Resident is not culpable, and I surely concur. If the facts were as he believes them to be, he is about to be killed by a CA. Quong then says that if resident is an IA and not a CA—as Resident surely is—then Resident is not liable to preventive force. I agree. Quong concludes, however, by arguing that the implication of Resident's being an IA rather than a CA is that, on my account, the twin brother may not use force to defend himself against Resident.

Quong is wrong, however. Even if Resident is an IA in attacking the twin brother, the latter may be an IA in defending himself. We may just have an ever possible tragic conflict between the two IAs. A TP, who knew all the facts, would surely see it this way. If he were to

[12] Id. at 53.

intervene, on whose side should he do so? Unless the TP could cite some moral tie-breaker, he would have no more reason to aid the twin brother than to aid Resident.

After rejecting other approaches to liability to preventive force, Quong offers his moral status account:

> A person is liable to defensive harm for choosing to do X when that choice results in a threat of harm to innocent people if and only if: (a) choosing X meets the minimum conditions of moral responsibility, and (b) the evidence-relative permissibility of choosing X depends either on the assumption that those who are harmed (or might foreseeably be harmed) by choosing X are liable to the harm, or else on false moral beliefs.[13]

The principal implication of Quong's moral status account is that some IAs *are* liable to preventive force, much as if they were CAs. For example, those IAs, like Resident, who threaten force against Vs because they nonculpably but mistakenly believe facts that would justify that force are, according to Quong, liable to preventive force in a way that other IAs—for example, the IA with the bomb-rigged cell phone—are not liable.

Suppose, then, that IA is a policeman. A has been seized by some evildoers and has been rigged up to look as if he is a suicide bomber. He is then released by his captors into a crowded mall. He begins screaming, and IA, believing A is a real suicide bomber about to detonate his explosives and kill several people, points a gun at him intending to shoot him and prevent the detonation. TP, who sees that the "bomb" is fake, may shoot IA to defend A. He

[13] Id. at 67–68.

need not, according to Quong, view this as an innocent versus innocent tragedy. Rather, if he intervenes, he must do so on A's side, not IA's. Because IA mistakenly assumed A was a CA and lacked the rights of the innocent, IA is liable to defensive force as if he were a CA even though his mistake was not culpable.[14]

Quong argues that mistakes about justifications that assume others lack certain rights are different from imposing justifiable risks on others even when, post hoc, the risks were unjustified (because 100 percent). I fail to see why, however. Consider some other cases of nonculpable mistakes. Alice is small and weak. She has a very ill but not mortally ill passenger whom she is driving to the hospital on a very narrow mountain road, with a cliff wall on one side and a sheer drop on the other. Alice sees a body lying in the road and which is impossible to avoid. She tries to move it to the side of the road, which would give her room to pass. It is too heavy, however. She finally decides that because it is almost surely—99.99 percent likely—a corpse, she will just drive over it. As she is about to do so, however, TP, who is far away, and who carries a rifle, sees what Alice is about to do. TP knows that the "corpse" is actually alive, and TP had been going to get help for him. TP shoots at Alice to prevent her driving over "corpse."

Would Quong say that TP was justified because Alice was liable to preventive force? She was, after all, assuming that "corpse" lacked rights. Of course, Quong could say, that Alice was not aware that she was risking harm to a living person, and that distinguishes her from the policeman IA's using force against A, the fake suicide bomber. It is not clear why taking a risk regarding another's rights potentially renders one liable to defensive force whereas taking a risk

[14] Id. at 64–65.

regarding whether someone is alive and has rights does not. (And compare Alice with Meg, the deer hunter, who nonculpably intends to shoot at a deer, but who is aware that occasionally a "deer" is, in fact, a hunter; on Quong's account, if the "deer" *is* a hunter, is Meg liable to defensive force as a CA would be?)

In sum, I believe the subtle distinctions Quong attempts to draw among the various IAs are unpersuasive. There are many ways IAs can be innocent, but they all add up to innocence, which does not come in degrees so far as I can see. The risk an IA takes that another lacks rights (because he is a CA or a corpse) seems no different from other risks that may turn out bad.[15]

* * * * *

My conclusion regarding IAs, then, is that neither McMahan nor Quong has produced a persuasive case for drawing distinctions among IAs and treating some IAs as if they were CAs. IAs are as innocent as Vs. From TP's perspective, siding with either IA or V is justifiable only on a lesser evils model; therefore, there must be a reason other than McMahan's or Quong's for siding with V. The numbers might do that (if the numbers count), so that if there are two Vs attacked by one IA, the TP can defend the Vs in order that the fewest people suffer harm. On the other hand, when there are two IAs and only one V, the TP would be compelled to refrain from using force against the IAs but might be justified in using preventive force against V if V is threatening preventive force against the IAs. Perhaps there will be reasons other than numbers for the TP to choose one side or the other. (Perhaps V is on the verge of a cancer cure but the two IAs are not.) And if the lesser evils model favors V, V can avail himself of that justification

[15] Ferzan concurs. *See* Ferzan, *supra* note 5, at 682 n. 49.

in using preventive force against IA; but V may also be able to invoke the excuse (or agent-relative permission) of duress if the lesser evils justification is not available. The critical point is that IAs are just as innocent as Vs and presumably have a moral status equal to that of Vs.

F. The Anticipated Innocent Aggressor

My next character is the anticipated innocent aggressor (AIA). An AIA is exactly like an ACA, except that the AIA is neither culpable nor anticipated to become culpable. Here is an example of an AIA:

> The dictator of X has told his inner circle that he intends to launch a nuclear missile against the U.S. at twelve noon. A C.I.A. mole in the inner circle relays that message to the president, who in turn sends a message to his Special Ops team on the ground in X to try to prevent the launch. X's missile silo is manned in shifts, and one shift ends at 11:45 a.m. Soldier is due to man the silo for the next shift and is on his way there. He is unaware of the dictator's decision and expects another boring and uneventful day in the silo. But he has been thoroughly brainwashed into believing that the dictator is highly moral and that the U.S. is evil, and if given the order to launch the missile, he will surely do so. At the moment, however, he has no inkling that he will be so ordered and thus has no intention to launch the missile. As Soldier approaches the silo, the Special Ops team is lying in wait. They intend to shoot him now, at 11:45, because once he is inside the

silo and receives the dictator's order, they will not be able to stop the launch.

Soldier is an AIA. If he launches, he will do so as an IA. That is why before he forms the intention to launch, he is an AIA. So the question is, then, whether the Special Ops team would be justified in shooting him. (Perhaps they do not need to be justified given the horrors of a nuclear strike; perhaps there is an analogue to duress that applies to national defense against WMDs.) I think the answer is that they are justified, and that killing Soldier is a lesser evil than the nuclear attack he will launch. After all, he is not being used as a means, assuming that deontological constraint is not itself overridden by the horror involved (which threshold deontologists might claim).

What if, however, the case of the AIA were like the example I gave of the ACA? In other words, what if there were only one potential victim of the AIA's anticipated attack? Would V or TP be justified in using preventive force?

Although V might be excused for doing so, neither V nor TP would be justified. For what is feared is that AIA will become an IA, and IAs are not liable to preventive force if such use does not fit the lesser evils model.

Finally, if one does not regard ACAs as truly culpable—because their latent culpability has not been actualized in a conscious culpable intention—then ACAs are just AIAs, and the analysis of AIAs applies to them as well.

G. Innocent Bystanders

Innocent bystanders, or IBs, including those involuntarily or unknowingly shielding Attacker, are those persons who are at risk of being injured by V's or TP's use of preventive force. If that use is to be justified, the risks of harm to IBs must be factored into the lesser evils calculus. Even if, however, the risks to IBs is so great that TP will not be justified in using preventive force, V may still have an excuse (or agent-relative permission) for doing so. In other words, risks to IBs may exceed the level for an agent-neutral justification but still not defeat V's claim of excuse (or agent-relative permission). (It is possible that V's excuse might even extend to using an IB as a shield.)

* * * * *

These are the characters, the dramatis personae, in the scenes of SD and DO. On the side of use of preventive force are V and TP. Imposing the threats to which they are responding are CA, CP, CF, ACA, IA, and AIA. In the line of fire is IB. But what kind of preventive force are V and TP *justified* in using against those characters who are liable to justified force? And what kind of preventive force is V *excused* in using when he is defending against one who is not liable to *justified* preventive force? That is the topic to which I now turn?

II. Proportionality and Its Corollaries, Necessity and Retreat

Proportionality, necessity, and retreat are typically regarded by both the doctrine and by theorists as three separate conditions limiting the permission to use defensive force. *Proportionality* is understood to require that the level of force employed defensively be roughly proportionate to the harm that the defender seeks to avert. Thus, deadly defensive force can only be used to avert death, serious bodily injury, rape and other similarly grave harms.

The *necessity* condition requires that the level of defensive force be no greater than that necessary to avert the threatened harm. Thus, if V or TP can ward off a deadly attack with his fists, he cannot use a gun. The *retreat* condition requires that if V can safely retreat from the threat, he must do so rather than use defensive force. (The retreat condition, unlike the proportionality and necessity conditions, does not apply to TP for the obvious reason that it is not the TP who is in danger.)

When properly analyzed, these three requirements turn out to be only one requirement—proportionality. To see this, consider that the proportionality requirement is really a requirement that V or TP must sacrifice lesser interests of his to avoid sacrificing greater interests of Attacker. Thus, if a wife can avoid getting slapped around by her husband only by using deadly force—she is not strong enough to avert the slapping by nonlethal means—then she must endure the slapping rather than take her husband's life.

The necessity requirement amounts to the same thing. Suppose V or TP can avert a lethal attack by using a gun (deadly force) or by wrestling Attacker to the ground. The latter defense will be effective, but it will require V or TP to expend great amounts of energy and to endure cuts and bruises. Using the gun requires little energy and will not injure V or TP in the slightest. If the necessity requirement rules out use of the gun, then it amounts to a requirement that V or TP sacrifice lesser interests of his—avoiding expenditure of great amounts of energy, avoiding cuts and bruises—in order to spare Attacker's greater interest (his life).

Note that there is another aspect to necessity that is neglected by most theorists. Consider that V or TP might have defensive options, but they may differ in likely effectiveness. Suppose V or TP could use his fists against an Attacker, which would be unlikely to kill or seriously injure Attacker, and avert a lethal outcome (for V) with a 25 percent likelihood. Or V (or TP) could use a baseball bat, which would be more likely to kill or seriously injure Attacker, and avert a lethal outcome to V with a 60 percent likelihood. Or V (or TP) could fire a gun, which would be extremely likely to kill or seriously injure Attacker, and avert a lethal outcome (to V) with a 90 percent likelihood. Finally, V or TP could throw a hand grenade at Attacker, which would kill Attacker and save V with almost a 100 percent likelihood. What force is *necessary* in such a case?[16]

It is now easy to see how the retreat requirement also fits with proportionality. For it again requires V to sacrifice a lesser interest—namely, to be in a place where he has a right to be—in order to avoid sacrificing Attacker's greater interest in his life. Only some jurisdictions have a retreat doctrine; nevertheless, it is a logical corollary of the more universal proportionality and necessity doctrines. And like effectiveness in the requirement of necessity, the "safely" condition on the requirement of retreat obviously is a matter of degree; no retreat will be guaranteed to be 100 percent safe. (Query: How does or should one distinguish a requirement of retreat from a requirement that one not venture out on pain of forfeiting nonculpable self-defense—that one not go "looking for trouble"? And when is the requirement of retreat triggered? When the train carrying the killers arrives at the station? When the killers' shadows show them to be near the street where one is standing? When they are on that street but beyond gun range? Obviously, retreat must occur at some time before it is too late to do so. But short of that, when is it premature? Note also that the time of retreat will

[16] Seth Lazar has a nice account of the relation of necessity to proportionality and of the relation of necessity to effectiveness. Seth Lazar, "Necessity in Self-Defense and War," 40 PHIL. & PUB. AFF. 3, 10–23 (2012). With respect to the latter relation, he proposes the following:

> *Necessity*: Defensive harm H is necessary to avert unjustified threat T if and only if . . . [the Defender judges] that there is no less harmful alternative, such that the marginal risk of morally weighted harm in H compared with that in the alternative is not justified by a countervailing marginal reduction in risked harm to the prospective victims of T.

Id. at 13.

affect how safe it is, as retreat when the killers are at the train station will be safer than retreat when they are just out of gun range.)

Typically, criminal law doctrine imposes these requirements in all instances where defensive deadly force might be employed. But it is a fair question whether, if Attacker is a CA, these requirements are warranted. For why should V be required to sacrifice *any* legitimate interest when faced with an Attacker who is culpable? (Notice that if V violates the proportionality, necessity, and retreat requirements, he becomes a CA vis-à-vis Attacker.) Consider the example above of one who faces being slapped around (but not seriously injured or killed) by a culpable assailant against whom she has no other defense than a lethal one. Why should she sacrifice her interest in bodily integrity and the avoidance of pain to spare the life of one who is culpably attacking her?

In other contexts, we can make committing even property crimes extremely dangerous for those who are culpable. With appropriate warnings we can protect our belongings with electric fences, razor wire, vicious dogs, deep moats, and other mechanisms that render theft dangerous to the thief's life and limb. Presumably, we are permitted to do things which would escalate another's minor culpable act into a quite dangerous one that would entitle us to use deadly force to defend ourselves (for example, V could lie down on a path on his property that a trespasser drives on, escalating the CA's trespassing into a threat to V's life and permitting V to use deadly force if CA does not stop his vehicle). If this is correct, why is the case different when CA is threatening V with a minor rights violation, such as the infliction of pain, and V has no opportunity to escalate CA's offense into a more dangerous one, as in the preceding example (perhaps by attaching a bomb to her body that would detonate if she is slapped)? In general, why should not V be able to avoid pain from a CA by a credible threat of lethal response? Theorists of self-defense need to answer this question.

It may be the case that this is another place where it is important to distinguish CAs from IAs. Proportionality and its corollaries, necessity and retreat, seem appropriate when Attacker is an IA. If V is excused for privileging his life over the life of an IA, that excuse should be a limited one and should not apply when V can, at a lesser cost than loss of *his* life, spare the life of the IA. And, to flag an issue that I shall take up in section IV, perhaps he should be quite confident that Attacker is a CA and not an IA before he can use disproportionate or unnecessary force or stand his ground rather than retreat. Indeed, perhaps the proportionality/necessity/retreat doctrine can be justified in *all* cases on the grounds that V can never know for certain that Attacker is a CA rather than an IA. Still, if there is a level of confidence that Attacker is a CA that is sufficiently high, in theory at least, V perhaps should not have to limit his response to one that is proportional and necessary. (Nor should Third Party for that matter.)[17]

[17] Joanna Mary Firth and Jonathan Quong provide one argument for a proportionality/necessity/retreat limitation on preventive force against CAs, one that is based on what they call the "humanitarian rights" of CAs. Humanitarian rights are rights to be provided with protection from serious harms when others can do so at a modest cost. They argue that CAs do not forfeit their humanitarian rights; thus, when defending against them with only proportional force is not terribly costly as compared with defense involving greater than proportional force, they have a right that only proportional force be employed. *See* Joanna Mary Firth and Jonathan Quong, "Necessity, Moral Liability, and Defensive Harm," 31 LAW & PHIL. 673, 693–701 (2012).

On Firth and Quong's account, if V (or TP) breaches CA's humanitarian right by employing greater than proportional force, V (or TP) is now a CA vis-a-vis the original CA, who is now a V. Now, the original CA may use preventive force against V (or TP), though it, too, must be proportional—assuming, that is, CA cannot credibly signal his termination of his original culpable attack on V. Firth and Quong suggest that because the duty correlative to the humanitarian right is not an onerous one, the amount of force that can be used to prevent its breach must be minimal. Id. at 699. Suppose, however, that V can defend against CA's raping her with her fists or with a gun. Using her fists is only mildly more costly to V than using the gun; therefore, she would violate CA's humanitarian right were she to use the gun. Now suppose she does attempt to use the gun, and CA has no defense against that breach of his humanitarian right except to use *his* gun. Would Firth and Quong say that CA would be wrong to defend himself against V's violation of his humanitarian right with the only means he has to save his life?

Indeed, the notion of proportional response to a CA is much more complex than a simple comparison of the potential harm to V with the amount of preventive force V or TP must use to avert the harm. Should the CA's level of culpability matter or only the amount of harm his act is feared to threaten? And should V's or TP's estimate of the risk of that harm matter or only the amount of harm if that risk is realized?

Consider the following scenarios:

> *Involuntary Russian Roulette*: Deborah likes to point guns at people and pull the trigger. She puts one bullet in a six-shooter and spins the cylinder. Then she puts the gun in a bin with nineteen identical guns and shakes the bin. She then reaches in and pulls out a gun, aims it at someone, and pulls the trigger. V (or TP) knows this, and today he finds Deborah pointing a gun at V, about to pull the trigger. May V (or TP) use his gun (deadly force) to prevent Deborah from pulling the trigger?

In *Involuntary Russian Roulette*, V is facing a one in one hundred and twenty chance of being killed if Deborah pulls the trigger, which is what Deborah intends to do. Deborah is a CA. V's use of deadly force is a proportional response to some level of threat to his life. Is it proportional to *this* level of threat from a CA?

> *Reckless Driving*: Don is driving recklessly. He is imposing a risk of colliding with V's car in order to make his tee time at the golf course. If he collides with V, he will most certainly damage V's car, and quite likely cause V to suffer cuts and bruises. There is a slight chance (as V or TP sees it) that the crash will kill V. Don is adverting to the risk of a collision but does not believe his driving is creating any risk of death. However, Don is a CA. May V (or TP) use deadly force to prevent Don's possibly colliding with V's car?

In *Reckless Driving*, V is facing some risk of death from a CA. The risk is a product of the risk of a collision times the risk that the collision will cause V's death. Assume the risk of death that V or TP estimates is the same as that in *Involuntary Russian Roulette*. Would the use of deadly force to avert the threat be proportionate?

Eggshell Skull: V has the proverbial eggshell skull. Doris is about to slap V for no reason sufficient to justify doing so. Ordinarily, such a slap would merely cause a bit of pain but no injury. However, a slap to V's head has a one percent chance of causing V's death. V has warned Doris not to slap him, but she does not believe his claim of an eggshell skull. As she draws back her hand, may V (or TP, who knows V's condition) use deadly force to avert the slap?

Doris is a CA. And her intended act causes a risk of killing V. However, she is surely minimally culpable, even less culpable than Don and much less culpable than Deborah. For unlike them, she does not believe her act carries *any* risk of death or risk of serious bodily injury.

If we assume the CA's mental state in all three of the above scenarios is transparent to V and TP—they know all three CAs *are* CAs, and they know that Deborah and Don but not Doris are conscious that they are imposing some risk of death (Deborah) or serious injury (Don)—then what preventive force by V and TP would be proportional and hence justifiable? If deadly force would be justifiable to stop Deborah from pulling the trigger, should it not also be justifiable to stop Don's reckless driving?

On the other hand, although Doris is both culpable and is risking V's death, she is not aware that she is risking V's death. Should that matter in gauging what is proportional response? One could surely argue that it should matter. Although V, in fear for his life, might be *excused* in using deadly force against Doris, neither he nor TP would be *justified* in doing so. Or at least, that is how the argument would go if one thought that the proportional force limitation on justified preventive force should apply to threats from CAs.

Quong denies that culpability should matter, and that all CAs should be treated identically with IAs who are nonculpably imposing similar unjustified risks on Vs. He bases his denial on the following argument:

> (1) If two CAs are imposing identical risks on V, but one CA is more culpable than the other, it is counterintuitive that V should be restricted to lesser force against the less culpable CA. Therefore, defensive force need not be proportional to culpability.
>
> (2) If, on the other hand, there is a culpability threshold, then if the threshold is low, one may use the same defensive force against a minimally culpable CA as one can against a maximally culpable CA imposing the same risk of harm; but if the threshold is set higher, two CAs, who differ only slightly in culpability, will be liable to quite different degrees of defensive force.[18]

I must confess that I find neither (1) nor (2) persuasive. If one does believe that the proportionality constraint applies to SD and DO against CAs, I see no reason to think that constraint should not be sensitive to degrees of culpability. Alternatively, proportionality might be thought to be coarse grained, in which case a threshold account might make sense. Ultimately, Quong rejects culpability as a basis for justified (non-excusable) preventive force,

[18] Quong, *supra* note 10, at 50–53.

and he draws his distinctions in a way that cuts across the CA-IA divide. I have already expressed my doubts about that approach.[19]

III. The Trigger for Use of Defensive Force: Act? Intention? Probability?

A. The Trigger for Excusable Defensive Force

If we focus on V and ask when V's use of defensive force might be excusable rather than justifiable, the principal focus will be on V's fears regarding Attacker (whether a CA or an IA) and V's alternatives to the use of defensive force. In section II, we discussed V's alternatives when we took up proportionality and its corollaries, necessity and retreat. The focus here will be V's level of fear and its basis.

V's level of fear turns on (1) V's assessment of the probability that Attacker will in the future act so as to impose a risk of harm or harms on V and (2) V's assessment of the level of that risk and the magnitude of the harms risked. Suppose, for example, that Deborah of Involuntary Russian Roulette does not know that if she pulls the trigger of the gun she is pointing at V, there is a 1 in 120 chance that V will be shot. She erroneously believes the guns are all toy guns and are not capable of firing bullets. She is just playing. In other words, she is an IA.

Suppose that V knows that the guns are real and that there is a live round in one of the twenty guns. So he realizes that if Deborah pulls the trigger, there is a 1 in 120 chance he will be shot. Suppose he believes that if he is shot, there is a good chance he will die and an even better chance he will be severely injured. Further, he believes that it is almost certain that Deborah will pull the trigger. Finally, he perceives no avenue for retreat and no means of

[19] *See* text at nn. 10–14 *supra*.

preventing Deborah from pulling the trigger other than by shooting her, which will likely kill her.

V's level of fear is generated by the probability that he will suffer serious injury or death unless he shoots Deborah. The probability is the product of the several probabilities involved: that Deborah will pull the trigger, that the gun has the live round in the firing chamber, that the bullet will strike V in a vital spot, and so forth. The question to ask if V shoots Deborah and then seeks to be excused is whether that ultimate probability times the magnitude of the harm is high enough so that a person of reasonable firmness would act as V acted.

Similarly, if V did have an avenue of possible retreat, but taking it would have reduced the probability of being shot and seriously injured or killed only slightly, then we should ask if this latter risk was still high enough to excuse V for shooting Deborah rather than trying to retreat. And the same goes for possible alternatives to shooting Deborah, such as trying to grab her gun or telling her the gun is real and might be loaded. If attempting to grab the gun would not reduce the risk appreciably, and if it is highly unlikely that she will believe V's assertion that the gun is real, the magnitude of the risk of taking these alternatives might still be sufficiently high to excuse V for shooting Deborah instead.

B. The Trigger for TP's Justified (Lesser Evils) Defensive Force Against IAs

If the TP believes Attacker is an IA, then TP's use of force to defend V must be justified by a lesser evils calculus. That calculus would require TP to estimate the risks of harms to the IA of TP's use of defensive force and compare that estimate to his estimate of the risks of harms that IA will impose on V if TP does not use defensive force. Only if the risks of harms facing V if TP does not act exceed the risks of harms to IA if TP uses preventive force will TP be justified in using preventive force as a choice of the lesser evil. Moreover, TP must also assess alternatives available to him—using lesser force against IA or using means other than force—to determine whether his use of preventive force against IA really is the choice of the lesser evil.

C. The Trigger for Use of Justified Preventive Force Against a CA: Probability, Act, or Intent?

Thus far we have discussed the triggers for excusable preventive force against any Attacker (IA or CA) and for justifiable preventive force against IAs. In both cases, the trigger turns out to be a probability of harm threshold in the former case and comparative probabilities of harm in the latter. The question now is whether justified preventive force against CAs—whether employed by V or TP—itself turns solely on probabilities, or whether instead it can be triggered by CA's acts or CA's intention. We implied earlier that harms to CAs can be discounted, perhaps severely, in a justification calculus, so that if three CAs are murderously attacking one V, TP would be justified in killing all three to save V's life. (Obviously, if TP would be justified in doing so, so would V.) If that is correct, then this must be because CA's culpability in some way triggers a discounting of his value in an agent-neutral justification calculus. The question then is, what exactly is it about CA that triggers the discounting, and does whatever it is supplant a consideration of the probability of harm to V or instead work in tandem with it?

Take, for example, the original Deborah playing involuntary Russian roulette on V. She is a CA, for she knows that she's imposing a 1 in 120 risk of death or serious injury on V if she pulls the trigger, which she now intends to do, and that there are no facts of which she is aware that would justify her doing so. Does her mere intention render her liable to justifiable preventive force, or must she take some action based on that intention? (Note that this question is different from the questions raised in section II, where we asked what preventive force would be a proportional response to Deborah's threat and whether a proportionality constraint should even apply to a CA like Deborah.)

We can easily rule out an act requirement. An act would be necessary to render one a CF[20]—the CF would need to do or at least say something to make V or TP believe CF was a CA—but an act hardly seems necessary in order to be a CA.

It seems obvious that the CA need take no specific act to be liable to justified SD or DO. Dinah may be picnicking on the edge of the cliff on which sits Balanced Rock when she sees the dynamite and detonator that Dan has left behind and sees her arch-enemy Victor walking in the valley below. She realizes that if she pushes the plunger on the detonator, the resulting explosion will unbalance Balanced Rock, resulting in its falling on Victor below. She decides to do that in order to kill Victor. At the moment she forms that intention, before she has engaged in any specific act of aggression, she has become a CA. For she is in the same position she would have been in had she formed the murderous intention earlier and brought the dynamite and detonator to Balanced Rock herself; and surely in that case she would have been a CA. So it is her intention to kill Victor in circumstances in which she is unaware of any facts that would justify her doing so that renders her a CA and liable to justified preventive force.

The question now is whether a culpable intention by itself renders one liable to justified preventive force or whether the culpable intention must be accompanied by a certain level of risk of harm. Recall that this risk is the product of the probability that the CA will in fact act as

[20] *See* Ferzan, *supra* note 5, at 693. If V or TP would be justified in using preventive force against a CA, they will not be culpable if the CA turns out not to be a CA but is instead a CF.

he now intends[21] and the probability that if he does, harm will occur. This latter probability is really a set of probabilities of different kinds and magnitudes of harms.

To simplify, let us focus on just one harm of a specific magnitude—death. Return to Deborah, playing involuntary Russian roulette on V. The relevant risk faced by V is, if we assume no God's-eye view of the situation, a 1 in 120 risk of V's death discounted by the chance that Deborah will not pull the trigger despite her present intention to do so. Suppose Deborah is 90 percent likely to pull the trigger, so that V faces a 1 in 132 risk of being killed. Is this risk sufficient to justify preventive force against Deborah by V or TP? Deadly preventive force?

Note that if V or TP perceived *no* risk that Deborah would kill V, perhaps because they thought her gun was a toy, their use of force against her would not be preventive force. They would be acting culpably, *even if we discover ex post that her gun was real and that V did face a risk of death*.

On the other hand, if V or TP do perceive a 1 in 132 risk that Deborah, a CA, will kill V, is that a sufficiently high risk to justify preventive force, and, if so, preventive force of what magnitude? This issue is the issue of proportionality that I discussed in section II. Are CAs subject to a proportionality constraint, and if so, how does it apply to Deborah, Don (the reckless driver), and Doris (unaware of V's eggshell skull)?

A second question here is whether there should be a probability threshold for the CA's acting as he intends, or whether there should just be a single probability threshold that is the product of the probability of the CA's acting as he intends times the probabilities of harms if he

[21] The probability that CA will act as he now intends is a function in part on the temporal period between the formation of the intention and the time its execution is to take place. The longer that period, the more things might occur that would cause the CA to revoke his intention or preclude his acting on it. Time, therefore, affects only the probability of harm faced by V. There should be no requirement for justified preventive force that the feared attack be "imminent" if its probability is otherwise high enough.

does so. If the CA intends to detonate an atomic bomb but is highly unlikely to do so, does this latter tiny probability preclude justified preventive force even if the overall probability of harm, given the magnitude, is quite high? (Assume the only chance to use preventive force is now, when the chance that the CA will act as he intends is quite low.)

Whatever the level of risk is that renders nonculpable preventive force against CAs, if V or TP perceive the risk to be above that level, their use of whatever degree of preventive force that is appropriate given the CA's culpability will be nonculpable even if ex post it turns out that there was no risk. If V or TP may nonculpably use deadly force against Deborah, the involuntary Russian roulette player, then they can do so nonculpably even if we later discover that there was no live bullet in the chamber. Whether an action is culpable is relative to the actor's epistemic perspective, even if one thinks *justified* action turns solely on how things really are. Thus, V's or TP's perception of the risk makes his act *nonculpable* even if, from a God's-eye or ex post perspective, he should not have used preventive force.[22]

IV. IA, CA, CF, or CP? Probability Regarding the Status of the Attacker and Others

Thus far we have stipulated the status of the Attacker—as an IA or CA—and the status of non-attackers who might nonetheless be nonculpably intentionally harmed—the CF and CP. But of course, the V or TP cannot know for certain whether the Attacker is an IA or a CA, whether the apparent Attacker is really not one but is a CF, or whether a bystander is a CP or an

[22] I first used the term "justifiable" in writing this paragraph in place of the term "nonculpable" that I now use. When someone acts based on an estimate of the probabilities of those factors that bear on its moral status, both terms—"justifiable" and "nonculpable"—seem inadequate. The former seems inadequate because if God or some more knowledgeable observer than V realizes there is no bullet in Deborah's gun, from that perspective, V's act of killing Deborah in SD is morally regrettable. The latter seems inadequate because it effaces the distinction between those defenders whose actions cannot be faulted and those whose actions are criticizable but excused. Perhaps we need terms like "nonculpable (J)" and "nonculpable (E)" in order to distinguish those who acted properly based on their probability estimates—even when, ex post, when the probabilities resolve into 1s and 0s, their acts turnout to have been morally regrettable—and those who acted wrongly but excusably.

IB. If the permissibility of TP's and V's use of preventive force and its magnitude turns on the status of the one against whom that force is used, how certain must TP and V be with respect to that status?

For suppose TP or V thinks Attacker is a CA with a 40 percent degree of confidence—must he then act as if Attacker *is* an IA? Theorists of self-defense tend to stipulate that Attacker is a CA or an IA. But in the real world, TP or V will believe Attacker is a CA or IA (or CP) with a level of confidence greater than zero but less than 100 percent.

Consider this example. TP or V knows that two sets of twins are out to kill V. One set, Pixie and Trixie are classic IAs. Perhaps they've been told by a (to them) credible source that V is a terrorist and must be killed on sight to avert a disaster. Or perhaps they're insane.

The other set of twins, Dina and Dana, are classic CAs. They just hate V and wish to do him in.

Suppose TP knows all this but does not know what the twins look like. He sees a pair approaching V with guns drawn. If they are Pixie and Trixie—IAs—then arguably it will be unjustifiable and hence culpable for TP to kill them in order to save V, no matter how high TP assesses the likelihood to be that if he does not intervene, V will be killed. For there is no obvious reason for TP to prefer the life of one V over two IAs. (V also would not be justified in saving himself at the cost of two IAs' lives; but V might have an excuse or agent-relative permission for killing them, which is why I focus here on TP.) On the other hand, if the twins are the CAs, Dina and Dana, killing them will be justifiable. Now suppose TP considers this and concludes that it is 50 percent likely that the twins are Pixie and Trixie and 50 percent likely that they are Dina and Dana. TP's only options are to shoot the twins or abstain. Abstention is permissible but will result in V's death, or so TP believes. Would shooting the twins also be permissible given TP's belief that it is 50 percent likely that they are IAs? Would *any* chance that they are IAs require TP to treat them as if they are IAs, in which case, because there is *always* a chance that an apparent CA is really an IA, TP—and V as well—would always have to act on the assumption that an Attacker is an IA. The question is one of the required confidence level in the material facts, and it is a question on which almost every self-defense theorist has been completely silent.[23]

The same confidence level problem comes up when TP sees V apparently threatened, not by a CA who *intends* to kill V, but by a reckless CA. That would be the case if the apparent CA is, for example, playing Russian roulette on V. But the apparent CA might be insane and thus an IA. And the same problem arises when TP sees apparent CA driving very fast on a narrow road that V is on, and TP can only protect V by shooting CA or shooting his tires (which will likely kill him by causing his car to go out of control and careen off the road and over a cliff). Apparent CA might be culpably reckless and thus an actual CA. But he also may not be a CA at all, as would be true if he has a very ill passenger he is rushing to the hospital, making the risks he is imposing on V justifiable.

It is clear then that for the use of preventive force either in SD or DO to be agent-neutrally justifiable (and not merely, in the case of SD, excusable or agent-relatively permissible), one must assign a required level of confidence to the judgment by TP or V that

[23] Kim Ferzan and I do make a first pass at this in our book, where we say that a TP, in doing the lesser evils analysis in a one on one confrontation in which V is being attacked, and it is uncertain whether Attacker is a CA or an IA, should use any probability that Attacker is a CA to discount Attacker's status and therefore intervene, if at all, only on V's side. LARRY ALEXANDER AND KIMBERLY KESSLER FERZAN, CRIME AND CULPABILITY: A THEORY OF CRIMINAL LAW 123–26 (2009). That is pretty much the extent of the distance we went in trying to deal with the issue of uncertainty regarding Attacker's status.

Attacker is CA. Obviously, that level has to be greater than 50 percent, for if it were below 50 percent, TP or V would not believe Attacker was a CA. But how much above 50 percent must it be? Just over 50 percent? Beyond all reasonable doubt? Beyond all doubt? If the requirement is beyond a reasonable doubt or higher, then TP and V must respond to Attacker as if Attacker is an IA if their confidence that Attacker is a CA is not at that level. (Similarly, if CPs may be discounted in a lesser evils calculation or may be treated as a means, how confident must TP or V be that a bystander is a CP rather than an IB?)

I have no answers to proffer for this problem of the requisite confidence levels for statuses. Nonetheless, no theory of preventive force that makes status distinctions will be complete if it fails to confront and satisfactorily resolve this question.

V. Provoking the Attacker

What are the rights and liabilities of one who has provoked another to employ preventive force? If the provoker is a CA or CF, and is now being attacked defensively by TP or V, may CA or CF use preventive force in return? Because, as to the CA or CF, the TP or V is an IA, the CA or CF would surely not be justified in using preventive force. However, because Vs might be excused or have an agent-relative permission to use preventive force against IAs, it seems possible that CAs and CFs might also be excused in defending themselves against preventive force by TPs and Vs.

Why is that not an outrageous conclusion, a reductio of my analysis? The answer is that if one notes that the CA and CF will already be highly culpable for having created the risk of just such a deadly conflict, their being excused for defending themselves does not mean excusing them for having created the excusing predicament. They are in the highest of moral and legal hot water whatever action they now take to defend themselves, especially if, like me, you do not believe that results affect culpability and retributive desert.

What about actors who provoke others into becoming attacking CAs in order then to use preventive force against them justifiably? Suppose V has provoked a culpable overreaction by a CA for the purpose of getting the CA to overreact. (Think of Charles Bronson in *Death Wish*,[24] posing as a vulnerable target walking through Central Park at night, hoping that he gets attacked in a way that would make deadly defensive force justifiable. Think of Jack Palance in *Shane*,[25] insulting a hotheaded farmer in order to provoke him to draw his gun so that Palance can kill him in self-defense. Or think of Kathy Bates in *Dolores Claiborne*,[26] giving her abusive husband alcohol and accusing him of abusing their daughter in order to provoke him into a murderous rage and chase her down a path on which she had constructed a deadly booby trap.) Should V be deemed nonculpable for defending himself against an overreaction to his provocation if his provocation anticipated, or even was meant to induce, that overreaction?

Although this is controversial, I believe the answer should be "yes." When a CA overreacts to a provocation and launches a deadly attack on V, V should be deemed justified in responding with deadly preventive force even if V anticipated the attack, and even if V hoped it would occur. These scenarios are closely analogous to "stings" and entrapments, where latent criminals are lured into committing criminal acts for which they are then arrested. And

[24] *Death Wish* (Paramount Pictures, 1974).
[25] *Shane* (Paramount Pictures, 1953).
[26] *Dolores Claiborne* (Castle Rock Entertainment, 1995).

whatever we think of stings and entrapments by the police, we do not criminalize them when engaged in by private citizens.[27]

VI. Miscellany: The Unknowingly Justified and the Innocent Mistake

What if V is being attacked by a CA but is unaware of this and attacks CA for some reason that would not justify the attack? This is the problem of the unknowingly justified actor. My position is relatively orthodox. Culpability is all that matters for me, not results. Therefore, one who acts believing the circumstances are such that were his belief correct, his act would not be justified, is a culpable actor even if the circumstances turn out to be otherwise and make his act objectively justifiable—that is, nonculpable if undertaken by one fully aware of the actual circumstances.

On the other hand, one whose mistaken beliefs about the circumstances are such that were they correct, his act would be justifiable, acts nonculpably. And if that act is one of preventive force, then we likely have a tragic confrontation between an IA and V, or between two IAs, each possibly believing the other is a CA. From a TP's perspective, where the TP is fully informed, unless the numbers or some qualitative factors dictate an agent-neutral reason to favor one side in this tragic affair over the other, the TP must stand aside if he cannot avert the tragedy other than by force equal to that the parties are employing. No morality or law can guarantee the absence of tragic confrontations due to mistakes about how the world really is.

VII. Restraining Attackers

[27] The private citizen has, of course, encouraged, obliquely, the target of the sting to act culpably and thus is complicit in the culpable act. Bronson, Palance, and Bates are thus complicit in the culpable attacks against them. In those cases, all the risks presented by those attacks are internalized by the solicitors—Bronson, et al—who are prepared to shift those risks but only to their culpable attackers.

What if we employ non-deadly preventive force against Attacker and avert the attack, with the result that we now have control over Attacker? What can and should we do now? If Attacker is an IA, then, if he is morally and legally irresponsible because of mental illness but remains dangerous, we can civilly commit him until he is no longer ill or no longer dangerous. If he was innocent because acting under duress, we should eliminate the source of his duress, at which point he will no longer need to be restrained because he will no longer be a danger. If he was innocent because of a mistake, we should apprise him of his mistake and let him go. The prescriptions for dealing with IAs once their attacks have been averted are fairly obvious.

So, too, is the prescription for CFs. CFs never were a danger, but they can be punished for culpably creating fear and the risk of unnecessary violence.

CAs, however, present a different challenge. Although they may, like CFs, be punishable for culpably creating fear and the risk of defensive violence, Ferzan and I reject holding them criminally liable for attempting the intended culpable attack.[28] Rather, we have argued that they can be preventively restrained for as long as they continue intending that attack.[29] But that's a subject that I shall not pursue here.

VIII. Conclusion

The purpose of this paper has been to provide an inventory of all the factors that a theory of preventive force must take into account. I have, in addition, stated my opinions about how many of those factors should be treated. Still, I have left some issues about them

[28] *See* Alexander and Ferzan, "Risk and Inchoate Crimes," *supra* note 2, at 106–19; "Danger," *supra* note 2, at 641–60.

[29] Alexander and Ferzan, "Danger," *supra* note 2, at 660–67.

unresolved. For example, I have not resolved how ACAs should be treated—as CAs or as IAs—although I am inclined toward the latter. Nor have I resolved to the point of being convinced whether the interests of CPs can be discounted in a lesser evils calculus and whether CPs may permissibly be used as means, though I am inclined to answer both questions affirmatively. With respect to CAs, I have left unresolved whether proportionality, necessity, and retreat should constrain those using preventive force against them. Nor have I resolved whether, in defending against an Attacker, the probability that the Attacker will attack and the probabilities that such an attack will harm V in various ways should be treated separately or combined into one overall probability of harm. Finally, I have not even begun to resolve the question of how confident must V or TP be that Attacker is a CA, or that a bystander is a CP, to justify treating them as such rather than as an IA or IB.

So there are many loose ends in what I have here presented. Still, I hope to have shown how complex a theory of preventive force must be and to have made a good start at producing such a theory.