1998

Self-Defense and Objectivity: A Reply to Judith Jarvis Thomson

Russell Christopher
russell-christopher@utulsa.edu

Self-Defense and Objectivity: A Reply to Judith Jarvis Thomson

Russell Christopher *

In *Self-Defense*,[1] Judith Jarvis Thomson sets out a theory of self-defense rendering permissible the use of defensive force against both culpable aggressors and morally innocent threats, but not against bystanders. The ostensible purpose of Thomson's theory, and that which has attracted the most criticism,[2] is justifying (not merely excusing) force against morally and juridically innocent, but causally harmful, threats. Thomson adopts a rights-based version of moral forfeiture which utilizes an objective[3] perspective.

---

* Research Scholar in the Faculty of Law, Columbia University School of Law. I wish to thank Anthony Dillof, George Fletcher, Kent Greenawalt, Maria Pagano and Joseph Raz for their criticisms of an earlier draft of this article.

1. Judith Jarvis Thomson, *Self-Defense*, 20 PHIL. & PUB. AFF. 283 (1991) [hereinafter *Self-Defense*].

2. *See* Larry Alexander, *Self-Defense, Justification, and Excuse*, 22 PHIL. & PUB. AFF. 53 (1993); Jeff McMahan, *Self-Defense and the Problem of the Innocent Attacker*, 104 ETHICS 252 (1994); Michael Otsuka, *Killing the Innocent in Self-Defense*, 23 PHIL. & PUB. AFF. 74 (1994).

3. My use of the term 'objective,' and later in the paper the term 'subjective,' both of which are extraordinarily ambiguous, is adapted from Thomson's distinction between the terms as found in: JUDITH JARVIS THOMSON, THE REALM OF RIGHTS 172-73 (1990) [hereinafter THE REALM OF RIGHTS]; Judith Jarvis Thomson, *Imposing Risks*, *in* RIGHTS, RESTITUTION, AND RISK 178-79 (William Parent ed., 1986) [hereinafter *Imposing Risks*]. Conduct is permissible, under an objective approach, based on what was, is or will be the case, without regard to any mental state of the actor. Conduct is permissible, under a subjective approach, based on an actor's belief or knowledge of what was, is or will be the case or by consideration of one or more mental states of an actor.

Some theories might be called objective in that they contain one or more objective components or criteria while also utilizing subjective criteria. See, for example, the theories discussed in section VIII. These same theories could variably be termed partially objective or even partially subjective. For the purposes of clarity in contrasting other theories with Thomson's theory I will term any theory that contains at least one subjective criterion as subjective or partially subjective.

Later I will distinguish between two possible types of objective accounts which Thomson might be employing. I will term these the strongly and weakly objective approaches. *See infra* section II.

Although commentators have disagreed with her conclusions about favoring morally and causally innocent actors over morally innocent, but causally harmful, actors, no one has taken issue with her underlying objective approach to self-defense.

The doctrine of self-defense faces an intriguing conceptual puzzle. We conceive of permissible self-defense as a response to a previous or the initial impermissible threat of harm. Yet in order for force in self-defense to be successful it must, in some cases, be employed prior to the fruition of the threat to which it is in response. Since the force in self-defense may actually occur prior to or prevent the actuality of the threatened harm, the force used in self-defense may easily resemble or be confused with the initial impermissible threat of harm from an objective perspective ignoring the beliefs, intentions and reasons of the actors. I will argue that Thomson's theory fails to successfully distinguish between what we intuit to be the impermissible aggression or threat and the permissible use of self-defense in response to the aggression or threat.

This reply to Thomson's theory of self-defense will particularly focus on the untenability of her objective account of permissible force. I will claim that a careful analysis of Thomson's approach yields conclusions diametrically opposite to those which she claims. Her objective approach, properly considered, not only justifies the threats posed by morally innocent, but causally harmful actors, but also the force used by culpable (evil) aggressors and unwittingly forfeits the life of morally and causally innocent victims. The seemingly most attractive revision to her theory, which might avoid these counterintuitive results, will be shown to result in, among other problems, a paradox in which what is permissible becomes impermissible. I will argue that only the inclusion of subjective criteria will make Thomson's

---

For general discussions of objectivity and subjectivity see KENT GREENAWALT, LAW AND OBJECTIVITY (1992); Thomas Nagel, *The Limits of Objectivity, in* 1 THE TANNER LECTURES ON HUMAN VALUES (Sterling M. McMurrin & Eric Asby eds., 1980).

theory tenable. More broadly, my remarks may suggest significant difficulties with any objective account of justifiable self-defense.

I.

Thomson develops and refines her account of permissible self-defense by considering six hypothetical cases.[4] In Villainous Aggressor,[5] a driver of a truck is trying to kill you by running you over; you can only save your life by blowing up the truck (and thereby killing the driver). Thomson offers the argument that lethal self-defense force is permissible in this instance because he is villainously aggressing against you and that unless you kill him he will kill you. Thomson next considers Innocent Aggressor[6] in which the same facts pertain except that the driver is morally blameless (he has been drugged). She concludes that the lack of fault of the Innocent Aggressor is irrelevant and that self-defense force is still permissible against him. Through this example, Thomson revises her principle explaining self-defense derived from the first case. Defensive force is now permissible against both drivers because they are aggressing against you and they will kill you unless you kill them. In Innocent Threat,[7] a fat man accidentally falls off a cliff and will land on you thereby killing you. You do not have time to move out of the way, but you do have time to shift an awning so that he is deflected away from you, thereby killing the fat man. That the fat man is not aggressing against you Thomson finds irrelevant and concludes that self-defense force is permissible against the fat man. Thomson further narrows her principle explaining why self-defense force is permissible in all three cases by saying that the person you kill will otherwise kill you.

---

4. Readers familiar with Thomson's theory may well choose to skip ahead to the next section.

5. *Self-Defense*, *supra* note 1, at 283-84.

6. *Id.* at 284-87.

7. *Id.* at 287-89.

By next considering three bystander cases,[8] Thomson suggests that the principle is still incomplete. In each case, you can save your life by killing a bystander. Yet defensive force is impermissible against the bystanders because the above principle will not be satisfied; it is not true that the bystanders will otherwise kill you. In considering whether it is always impermissible to use defensive force against bystanders Thomson confronts the case of wartime bombing of civilians[9] (i.e., bystanders) and the Doctrine of Double Effect.[10] Although Thomson bypasses the issue of the permissibility of such force in wartime, she rejects the doctrine and advances two theses:

> 1. The Irrelevance-of-Intention-to-Permissibility Thesis: it is irrelevant to the question whether X may do A what intention X would do A with if he or she did it.
>
> 2. The Irrelevance-of-Fault-to-Permissibility Thesis: it is irrelevant to the question whether X may do A whether X would be at fault in doing it.[11]

Thomson illustrates her theses with the example of Alfred,[12] who has a dying wife, and buys what he believes is poison intending to give it to her to accelerate her death. Unbeknownst to Alfred, however, that which he believes is poison is, in fact, the only existing cure for his wife's malady. Thomson concludes that it would be absurd to claim that because Alfred would give it to her with a bad intention and that he

---

8. *Id.* at 289-92.
9. *Id.* at 292-93, 296-98.
10. *Id.* at 292-296. Thomson explains that:
> the Doctrine of Double Effect says that we may do what will cause a bad outcome in order to cause a good outcome if and only if (1) the good is in appropriate proportion to the bad *and* (2) we do not intend the bad outcome as our means to the good outcome.

Id. at 292.
11. *Id.* at 294-95.
12. *Id.* at 293-95.

would be at fault that it would be impermissible for him to give her the stuff which, in fact, will cure her.

Returning to the issue of bystanders, Thomson asserts that "[o]ther things being equal, every person Y has a right against X that X not kill Y."[13] Can Y's right be overridden? That X is at risk of death and can only save her life by killing bystander Y does not, Thomson concludes, override Y's right to life. This explains why bystanders cannot be killed. Yet every person's right to life can be overridden if they are "about to" violate another's right to life.[14] By being about to violate your right to life the Villainous Aggressor, for example, forfeits or loses his right to life. Because you can only save your life by killing the driver and he has forfeited his right to life, self-defense force against the driver is permissible.

Similarly, the Innocent Aggressor and Innocent Threat lose their right to life because they are "about to" violate your right to life. Since you can only save your life by killing them and they have lost their right to life, self-defense force against them is permissible. You are not violating their right to life because they have already lost or forfeited it by being "about to" violate your right to life. To the initial principle "they will otherwise kill you" Thomson adds "that if they kill you they will violate your right that they not do so."[15] The insufficiency of the initial principle Thomson claims to demonstrate by considering whether one of the aggressors or threats discussed above may fight back in permissible self-defense against you.

Thomson supposes that both Alice and Villainous Aggressor have antitank guns.[16] Alice may permissibly blow up the aggressor's truck. May the aggressor permissibly use his gun against Alice in self-defense? The aggressor might claim that he had a right that Alice not kill him and that if he did not kill Alice, Alice would kill him and violate his

13. *Id.* at 299.
14. *Id.* at 301.
15. *Id.* at 303.
16. *Id.* at 304-05.

right not to be killed. By being about to kill the aggressor, Alice lost or forfeited her right to life and the aggressor killing Alice would not violate her right to life. Thomson finds the aggressor's claim absurd because the aggressor lost his right to life by driving the truck at Alice. In other words, since the aggressor is the first party (between Alice and Villainous Aggressor) to be about to violate the other's right to life Villainous Aggressor is the one that forfeits his right to life. Any force that Alice uses against Villainous Aggressor is permissible since the aggressor has already lost his right to life. Alice's force does not violate the aggressor's right to life and thus does not trigger the loss of her own right to life. Consequently, force used by the aggressor against Alice's force does violate Alice's right to life and is thereby impermissible.

Under Thomson's unrevised principle of "they will otherwise kill you," however, the aggressor's force would be permissible against Alice's force. The aggressor could properly claim that Alice would otherwise kill him. Under the revised principle of "they will otherwise kill you and that if they kill you they will violate your right that they not do so," Thomson claims that the aggressor's force is "obviously impermissible."[17]

In section IV, I will argue that Thomson's approach can be shown to yield the opposite conclusion. It is the aggressor's force (e.g., Villainous Aggressor's) which is permissible and it is the defender's (e.g., Alice's) force which is impermissible. The Villainous Aggressor's "absurd" claim of permissible self-defense against Alice will be shown, under Thomson's objective theory, to be valid. Before I make a case for the Villainous Aggressor's claim, I will next briefly contrast objective from subjective theories and explicate Thomson's particular form of objective approach.

17. *Id.* at 304.

II.

Thomson appears to employ what might be termed a strongly objective account of permissible self-defense. An actor's force neither loses permissibility because of a bad intention nor gains permissibility because of a good intention. According to Thomson's theses, discussed above, the fault, intention or belief of an actor in using force is simply irrelevant to the permissibility of her conduct. Thus the husband's intention to poison his wife (with a substance that unbeknownst to him will save her life) or his belief that he is giving her poison does not bar the permissibility of his act. Similarly, if the husband had intended to give his wife what he thought was a cure, but was in fact poison, the husband's good intention would not make his conduct permissible or justified. The permissibility of the husband's act depends not on what his intentions are or what he believes (even if reasonably), but what has objectively occurred and what will, in fact, occur. Under a strongly objective theory, either the substance is a poison (making his conduct impermissible) or the substance is a cure (making his conduct permissible).

Applying this approach to self-defense, an actor's defensive force is eligible to be justified if she is, in fact, about to be harmed and if defensive force is, in fact, necessary. The appearance or reasonable belief of harm is insufficient. Just as the indication, appearance or reasonable belief that the substance was a cure would not make the husband's act permissible if the substance was, in fact, poison, so also the appearance or reasonable belief that an actor is about to be killed does not trigger permissible self-defense force if the apparent attack was not, in fact, about to occur. Furthermore, the lack of an appearance, indication or reasonable belief that an actor is about to be harmed is irrelevant if, in fact, the actor is about to be harmed and force in self-defense is, in fact, necessary. Just as the lack of appearance, indication or reasonable belief that the substance was a poison would not make the husband's act permissible if

the substance was, in fact, poison, so also the lack of appearance, indication or reasonable belief both that an actor is about to be killed and that self-defense by the actor is necessary does not prevent self-defense force from being permissible if the non-apparent attack was, in fact, about to occur and self-defense was, in fact, necessary.

In assessing the permissibility of an actor's conduct, the actor's conduct is not viewed from the actor's perspective but from God's or an omniscient being's eye view (an ideal observer) i.e., what actually occurred or what would have occurred. In determining which actor's force in a physical conflict is permissible or justified in self-defense, the central issue is which party was, in fact, the first to be about to use force i.e., the first to violate another's right to life.[18]

The strongly objective account discussed above should be contrasted with what I shall term the weakly objective theory. The weakly objective theory determines the permissibility of conduct based on probability. Thomson develops this approach in two previous works.[19] Conduct is impermissible if there is a sufficiently high probability that some proscribed result will occur; conduct is permissible if there is not such a sufficiently high probability. Adapting this approach to self-defense, A's force in self-defense against B is permissible if there is a sufficiently high probability that

---

18. For further evidence of Thomson's reliance on an objective approach, see Alexander, *supra* note 2, at 60 n.11:

> Thomson . . . objects to my asserting a right not to be killed culpably, where culpability in turn rests upon awareness that the killing is wrongful. She argues instead that rights are reducible to what we ought to do, and that what we ought to do does not turn at all on our beliefs.

For Thomson's endorsement of objective moral theory and a rejection of subjective moral theory see THE REALM OF RIGHTS, *supra* note 3, at 172-73, 175, 233-34, 241-42; and *Imposing Risks*, *supra* note 3, at 179.

Although in *Self-Defense*, *supra* note 1, at 310 n.18, Thomson disavows a portion of her discussion of self-defense in THE REALM OF RIGHTS, *supra* note 3, at 348-73, in regard to the permissibility of killing some to save more, she does not disavow any of the parts of *The Realm of Rights* from which I cite.

19. THE REALM OF RIGHTS, *supra* note 3, at 170-74; *Imposing Risks*, *supra* note 3.

B is the first (between A and B) to be about to kill i.e., the first to be about to violate the other's right to life.

Thomson points out that although taking probabilities into account entails a loss of some objectivity it nonetheless does not involve a slide into a subjective view.[20] Thomson distinguishes between objective probabilities and an "agent's estimates of probabilities."[21] By considering only objective probabilities and not the subjective assessments of probabilities by an actor, considering the probability of a harm ensuing from a threat or aggressor cannot be said to be subjective.

In *Self-Defense* Thomson appears to be employing the strongly objective approach. No examples are discussed in which the aggression or threat posed is uncertain and a matter of probability. Thomson declares that she will "leave open what should be said in cases in which it is not certain that the aggressor will cause you a harm if you do not kill him but only more or less probable that he will."[22] Since Thomson leaves to the side cases of uncertain harm and the cases I will discuss only involve certain harms, Thomson's possible endorsement of a weakly objective theory that takes into account probabilities should not apply. In any event, I will assume that Thomson's theory of self-defense, as developed in *Self-Defense*, is intended to be strongly objective. In section IV, I will attempt to support the claim that Thomson's strongly objective theory of self-defense yields radically counterintuitive results.

Both objective theories may be contrasted with various (at least partially) subjective approaches to permissible self-defense.[23] A purely subjective theory of self-defense ignores the actual circumstances but focuses on the actors' beliefs and intentions. For example, force used by an actor who honestly, but mistakenly, believes he is being impermissi-

---

20. *Imposing Risks*, supra note 3, at 187.

21. THE REALM OF RIGHTS, *supra* note 3, at 173 n.7.

22. *Self-Defense*, *supra* note 1, at 286.

23. For a fuller discussion of various (at least partially) subjective theories, see *infra* section VIII.

bly attacked is still eligible to be justified. Another common type of subjective theory (though not purely subjective) imposes an objective standard on the actor's belief, not merely requiring that it be honestly held but also that it be reasonable. A third type requires not only an actor's belief in, or knowledge of, the justificatory circumstances but also requires that the threat be actual rather than mistakenly perceived. A fourth type focuses not on the beliefs or intentions of the actor asserting a claim of self-defense but on the aggressor or threat. Defensive force is only permissible against a threat or aggressor who believes or intends to act wrongfully (a culpable or villainous aggressor or threat). To adopt Jeff McMahan's distinction,[24] the first two at least partially subjective theories may be termed "agent-centered" in that they focus on the mental states of the defender asserting a self-defense justification;[25] the fourth theory may be termed "target-centered" in that it focuses on aspects of the attacker. The third theory is both partially agent-centered and partially target-centered since it requires both an actual threat of harm by the aggressor or threat and a particular mental state of the defender.

III.

To see how Thomson's account can yield diametrically opposite results than what she claims let us first carefully consider Robert Nozick's original case of the Innocent Threat.[26] Innocent Threat (IT) is pushed off a cliff such that he will land on and kill an innocent (for the purposes of clarity we will hereafter refer to him as Vic) below. The situation is such that the only way Vic can save his life is by disintegrating IT with a ray gun. Under Thomson's account,

---

24. McMahan, *supra* note 2, at 268.

25. Another term which has been offered for such theories is "agent-perspectival." Suzanne Uniacke credits Robert Young for suggesting the term. SUZANNE UNIACKE, PERMISSIBLE KILLING: THE SELF-DEFENSE JUSTIFICATION OF HOMICIDE 17 n.6 (1994).

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

IT is the first (between IT and Vic) to be about to kill the other and thus the first to be about to violate the other's right to life. By being the first to be about to violate the right to life of the other, IT forfeits his right to life. Once IT loses his right to life, Vic's defensive force does not violate IT's right to life and since Vic will otherwise be killed, Vic may permissibly vaporize IT with a ray gun.

Although actor A being about to kill actor B would normally forfeit A's right to life, that would not be so if B had already lost her right to life by already being about to kill A. The key issue, therefore, is to determine which actor is the first to be about to kill the other (i.e., the first to be about to violate the other's right to life). I will argue that under a strongly objective theory, the first actor (between two actors) to apply lethal force (or who would have been if not stopped) is the first actor who is about to kill.[27] Thus if Vic kills IT before IT can land on and kill Vic, which Thomson believes Vic may permissibly do, it will be Vic who was the first (between IT and Vic) to be about to kill the other and Vic who is the first to be about to violate IT's right to life. As a result, it is Vic who has already lost his right to life by the time that IT is about to kill Vic. Thus, contrary to what Thomson claims, IT's force is permissible and Vic's impermissible under Thomson's theory.

To see how this curious and seemingly counterintuitive claim might be true let us carefully establish a timeline of the events in Innocent Threat. Under Thomson's theory, A does not have to actually kill B (who has a right to life) for A's right to life to be forfeited, but merely be about to kill B.[28] Were this not so, one might have to be already dead before one could permissibly use defensive force to prevent one's death which would be absurd. Thus Thomson properly allows the permissibility of defensive force from the time when another is about to kill one (who possesses a right to life). For the purposes of establishing our timeline, let us arbitrarily set the time period in which one is about to kill

27. I will defend this principle in Section V.

28. *Self-Defense*, *supra* note 1, at 301.

another as ten units of time before one would actually kill another (it could be 1, 5 or 15 units of time, seconds, fractions of seconds or whatever time period Thomson might accept). Thus, for example, starting at ten units of time before A would kill B (who has a right to life), A may be said to be about to kill B and thus starting at that point A loses his right to life.

The time period of ten units of time is purposefully left indefinite[29] so as to be compatible with whatever time period Thomson or the reader finds acceptable. Some controversy exists regarding when defensive force may be applied against a threat or aggression. Various standards include when the threat or aggression is immediate, when its imminent, when defensive force is immediately necessary or when its necessary.[30] Thomson's standard of about to provides little guidance but it may be safely assumed that Thomson neither wishes to justify force in self-defense before the point in time that it is necessary nor so late in time that the defensive force cannot be applied in time to prevent the threat or aggression. The indefinite time period of ten units should be similarly construed and should correspond with Thomson's time period of "about to."

Suppose IT is pushed off the cliff at Time Zero (T0) and he will land on Vic at T10 unless Vic uses defensive force before impact at T10. Given our stipulated time period of ten units of time, IT is about to kill Vic at T0. If IT is the first (between IT and Vic) to be about to kill the other, IT forfeits his right to life starting at T0. Whoever is the first to be about to kill the other forfeits his own right to life and is subject to permissible self-defense force from the other. If IT is about to kill Vic at T0, when is Vic about to kill IT? If

---

29. By leaving the unit of time undefined, the time period of ten units of time is also indefinite. Yet by stipulating the number of units of time as ten, the time period is sufficiently quantified for the purposes of comparing when each actor in a hypothetical may be said to be about to kill while still remaining indefinite.

30. For discussion of these standards and others see Richard Rosen, *On Self-Defense, Imminence, and Women Who Kill Their Batterers*, 71 N.C. L. REV. 371 (1993); Robert Schopp et al., *Battered Woman Syndrome, Expert Testimony, and the Distinction Between Justification and Excuse*, 1994 U. ILL. L. REV. 45 (1994).

Vic is to save himself, which Thomson believes is permissible, Vic must use defensive force at some time before impact (T10). Suppose Vic vaporizes IT with his ray gun at T5. That would be within the time period of when IT is about to kill Vic and it also saves Vic's life. Yet if Vic vaporizes IT at T5, the time period in which Vic is about to kill IT starts at Time Minus Five (T-5). Since Vic was about to kill IT prior to IT being about to kill Vic (T-5 occurs prior to T0), Vic is the first to be about to violate IT's right to life and Vic consequently forfeits his right to life. IT's conduct does not violate Vic's right to life because Vic has already forfeited it at T-5. Since Vic will otherwise kill IT, IT killing Vic, or being about to kill Vic, is permissible under Thomson's theory of self-defense. By carefully considering the timeline, Thomson's theory yields the opposite conclusion than which she claimed.

I do not mean to suggest by my argument, as critics of Thomson's theory have contended, that self-defense force against moral innocents like IT is impermissible. I will argue that it is the strongly objective perspective underlying Thomson's theory which is untenable. In the next section I will demonstrate that under Thomson's theory, contrary to Thomson's conclusion and our nearly undisputed intuitions, self-defense force is impermissible against a villainous aggressor but permissible against a moral innocent.

IV.

Let us consider the case of Villainous Aggressor, discussed above, in which a driver of a truck (hereafter VA) is villainously trying to run over and kill an innocent (again, for purposes of clarity let us call him Vic). The only way that innocent Vic can save his life is by blowing up the truck before it hits him, thereby killing VA. Thomson explains that force by Vic is permissible because VA is the first (between VA and Vic) who is about to kill the other, by driving the truck at him, thereby being about to violate Vic's right to life. VA thereby loses his right to life. Force used against VA does not violate his right to life because he

has already lost it. Vic using force against VA does not forfeit Vic's right to life because he is not violating VA's right to life. Since VA will otherwise kill him, Vic killing VA is permissible self-defense force.

Unlike the case of Innocent Threat, perhaps no one would claim that self-defense force is impermissible against a culpable aggressor such as VA. Moreover, no one would assert that innocent Vic has forfeited his right to life and that his defensive force is impermissible. Yet under Thomson's strongly objective conception of self-defense, contrary to what she claims, it can be shown that VA's force is permissible in self-defense against Vic's force and that Vic's force is impermissible. Let us again suppose that the time period of about to kill starts at ten units of time before lethal force is applied. Further suppose that at T0, VA starts to drive at Vic to kill him. Unless Vic blows up the truck, VA will run over Vic at T10, thereby killing him. Thus starting at T0, VA is about to kill Vic. Whether VA being about to kill Vic at T0 thereby forfeits VA's right to life depends on who (between VA and Vic) is the first to be about to kill the other. If Vic is to save his life, as Thomson and presumably everyone else thinks is permissible, Vic must blow up the truck and kill VA sometime between T0 and T10. If we suppose that Vic kills VA at T5, then Vic was about to kill VA at T-5. Since Vic was the first to be about to kill the other, Vic at T-5 is about to violate VA's right to life. By being about to violate VA's right to life, Vic forfeits his own right to life at T-5. VA cannot be about to violate Vic's right to life at T0 because Vic has already lost it. Force employed by Vic is impermissible since it would violate VA's right to life. Since Vic has already lost his right to life, and Vic will otherwise kill VA, VA running over and killing Vic (or at least trying to) is permissible self-defense under Thomson's theory.

Thomson's theory, through careful attention to the timeline, is shown to yield the opposite conclusion than what it purports to do both in cases where Vic faces an innocent threat and where he faces a villainous aggressor.

This would seem to show that the difficulty in Thomson's account is not justifying force against moral innocents. That Thomson's theory can justify self-defense force by a villainous aggressor against what would seem to be an innocent faultless person is especially troubling since virtually no one would claim that a villainous aggressor's force should be justified in such a paradigm situation of permissible self-defense. I will argue that the problem stems from employing a strongly objective conception of permissible self-defense. My claim will be buttressed by showing, in section VIII, that various (at least partially) subjective accounts of self-defense avoid the counterintuitive results of Thomson's theory. In the next section I will address some possible criticisms of the method I have thus far employed.

V.

My interpretation or use of Thomson's phrase "about to" may strike many readers as unusual or even counterintuitive. Many may feel one may properly speak of IT and VA being about to use force, as Thomson uses the phrase, but something is amiss when we say that Vic is about to use force at the point I claim. Some may find it strange to say that Vic is about to use force at a point when he neither intended to use force nor was aware that he would use force nor was even aware of a threat that might require defensive force. One might claim that about to does not strictly refer to a temporal condition or is not limited to referring to a time period. Even if Thomson intended a strictly temporal meaning of about to, it could be argued that her theory need not rest on that meaning and may be easily altered without losing any explanatory power.

Although I concede that my interpretation seems unusual at the outset, I will argue that it is nonetheless a plausible and correct interpretation in light of Thomson's theory. The extent to which it seems strange may attest to the enduring appeal of a subjective conception of self-defense, even to those who favor a strongly objective ac-

count. I assert that my interpretation of about to is entailed by a strongly objective conception of justification, but is incompatible with a subjective view. The absurd consequences (justifying self-defense force by VA and forfeiting the life of innocent Vic) derived from Thomson's theory may only be satisfactorily avoided, I will argue, by adopting some form of subjective approach.

A strongly objective view focuses not on the intentions, beliefs or perceptions of any of the agents involved but is concerned with what, in fact, actually happened and what, in fact, will happen. It adopts an ideal observer's eye view. The permissibility of the husband's conduct, in Thomson's example, is assessed by what, in fact, was the case and ignores his intentions, ignorance, mistaken beliefs and perspective. That Vic did not intend to use force at the time when I established that he was about to use force is irrelevant under Thomson's Irrelevance of Intention to Permissibility thesis. That Vic is not at fault for either the villainous attack or innocent threat and that VA, if not IT, is at fault in initiating the need for force are all irrelevant under Thomson's Irrelevance of Fault to Permissibility thesis. That Vic was not committing any act or engaging in any physical conduct that might give the appearance or indication that he was about to kill VA at the point I claim is irrelevant since Vic was, in fact, imminently about to kill VA and force by VA against Vic was, in fact, necessary. Under Thomson's strongly objective theory, the lack of appearance or indication that another is about to kill another is irrelevant if the actor is, in fact, about to kill another. The strangeness in saying that Vic was the first to be about to use force, despite our sense that he did not initiate either conflict, stems from the still remaining residue of using fault, intention or belief as criteria of permissibility. From the perspective of Vic and from the perspective of appearances Vic was not about to use force at the time I established, but from the perspective of an omniscient God's eye view it is entirely proper to say that Vic was about to use

force despite giving no sign that he would, his unawareness that he would and his ignorance of an impending threat.

I believe that what underlies our traditional assumption that someone like VA is the first to be about to use force is that VA is the first actor who we have some evidence for saying is about to use force.[31] Although Vic is temporally closer to using force than VA, we have grounds at an earlier point in time for believing that VA is about to use force than we have warrant for believing that Vic is about to use force. But the actor who we first have grounds for saying is about to use force is not necessarily the first actor who is, in fact, about to use force. Under a strongly objective approach, the standard is not the first actor we may have warrant for believing is about to use force, but the first actor who is, in fact, about to use force.

The following examples may alleviate some of the strangeness of my interpretation of "about to," under a strongly objective approach, as applied in the cases of Villainous Aggressor and Nozick's Innocent Threat.[32] Suppose we are watching a hockey game and during a brief timeout a fan confidently says that one of the teams, the New York Rangers, is about to score. You might say to the fan that his belief is unfounded considering there is no indication or physical manifestation that anyone is about to score (play has been briefly suspended for the duration of the timeout). Yet if the Rangers, shortly after the conclusion of the timeout, do suddenly score, you might claim that his assertion was not epistemically justified or warranted while still admitting that ontologically what occurred indicates that

31. Self-defense scenarios are typically set out as follows: A is about to kill V and will do so unless stopped; V kills A. But suppose we alter the way the same scenario is presented: V is about to kill A and will do so unless stopped; V kills A before A can use force to save herself. In the former depiction of the scenario we naturally assume that A is the aggressor and V the defender. But in the latter telling our intuitions may be unsettled because we no longer have the narrative cue that the aggressor is the actor who we first become aware will use force. In the latter presentation of the scenario we are less sure, I believe, as to which actor is the wrongful aggressor and which is the innocent defender.

32. *See* NOZICK, *supra* note 26.

his statement was true. A strongly objective theory is not concerned with whether a well-founded reason exists for believing that another is about to use force, but whether the other, in fact, was about to use force. Thus merely because there was no indication, appearance or epistemic basis for stating, at T-5, that Vic, at T-5, is about to kill VA or IT does not preclude Vic from being about to kill VA or IT at T-5 under a strongly objective approach.

Let us consider another example. Suppose that bomb A will explode at T5 and that bomb B will explode at T10.[33] Further suppose that we arbitrarily define when a bomb is about to explode as ten units of time before it actually does explode. If bomb A explodes at T5, then bomb A was about to explode at T-5. When bomb A explodes at T5 and bomb B has still not yet exploded, we may safely conclude that the soonest bomb B may be said to be about to explode is some time after T-5. Thus bomb A is the first (between bombs A and B) to be about to explode. It is clear that for any two bombs, the first bomb which, in fact, does explode is *necessarily* the first bomb that is about to explode. Suppose we do not arbitrarily define when a bomb is about to explode as ten unspecified units of time before it actually does explode. Even so, if bomb A explodes before bomb B explodes, bomb A is nonetheless the first bomb to be about to explode.

May the above rule regarding which of two bombs is the first to be about to explode be extended to determinations of which of two human actors is the first to be about to kill? Although there are certainly a myriad of differences between bombs and human beings, under Thomson's strongly objective theory a number of the differences are treated as irrelevant. An actor's intentions, beliefs and fault (or lack thereof) are declared irrelevant. On what basis can the above rule regarding which of two bombs is the first to be about to explode be accepted while still denying that Vic, at T-5, is the first to be about to kill IT and VA and thus the

33. Bomb B will explode at T10 unless the shock waves emanating from bomb A exploding disrupt the triggering mechanism of bomb B thereby preventing bomb B from exploding.

first to be about to violate the others' right to life? Just as the first bomb to, in fact, explode is necessarily the first bomb which is about to explode, the first actor who does, in fact, apply force is necessarily the first to be about to apply force under a strongly objective theory. Perhaps about to does not imply a temporal condition. Yet if about to does not refer to a time period, however unquantified, what does it refer to? Even if the phrase involves a time period perhaps it involves something else as well. Maybe it cannot be said that an actor is about to use force until he intends to use force or is aware that he will use force. But this interpretation involves a slide back into a subjective account and is contradicted by Thomson's principle of the irrelevance of intention. Perhaps an actor is not about to use force until there is a physical appearance or manifestation that he will use force. But appearances may be deceiving and are not relevant to a theory which assesses the permissibility of self-defense on what, in fact, occurred and not on what appeared to be occurring.

Even if the phrase "about to," under a strongly objective theory, does have a strict temporal meaning, perhaps Thomson's theory could avoid the absurd consequences of justifying the force of villainous aggressors by scrapping the about to construct without falling back into a subjective approach. Instead of being about to violate another's right to life and forfeiting one's own by being the first to be about to kill, perhaps the new principle could be the first to actually kill. Yet that standard would produce the same (presumably) undesirable result of forfeiting, e.g., Vic's right to life if Vic kills first.

Perhaps Thomson's theory could be amended such that the first actor to exhibit an actual (not mistakenly perceived) physical manifestation of being about to kill another is the first to be about to violate the other's right to life and thereby forfeits her own right to life. Both VA and IT exhibited a physical manifestation of being about to kill Vic by driving at Vic and falling toward Vic, respectively, prior to Vic exhibiting a physical manifestation of being about to

blow up VA and vaporizing IT. This amendment would seem to yield the permissibility of Vic's force against IT and VA, as Thomson intended, as well as avoiding the clearly undesirable result of justifying VA's force against Vic. Let us assume for the moment that a physical manifestation requirement (PMR) does successfully prevent the counter-intuitive result of justifying VA's aggression and forfeiting the right to life of innocent Vic. I will argue that it nonetheless raises substantial difficulties for a strongly objective theory sufficient to suggest the preferability of a subjective approach. I will then argue that incorporating the PMR into a strongly objective theory yields a paradox that may prevent the PMR from preventing the counterintuitive results.

Requiring a physical manifestation seems more suited to a subjective theory and is incompatible with a strongly objective theory. A PMR is plausible within a subjective theory as evidence that an actor could honestly and/or reasonably believe that self-defense force was necessary. Considering who first exhibited an external threat of physical harm is useful to assess whether an actor's belief in an imminent threat of harm is honestly held and/or reasonable. A perceivable threat of harm is more persuasive than a gut instinct to establish that an actor honestly believes, reasonably believes or knows that he is about to be attacked. Without some physical manifestation of an imminent attack or threat we might plausibly doubt that the actor intended to use defensive force rather than aggressive force. A prior physical manifestation might also serve to identify who was at fault in initiating the need for the actor to use self-defense force.

What would be the purpose of a PMR within a strongly objective theory? Unlike within a subjective approach, the actor is not required to act in response to, or because of, or believe that force is justified because of a perceived physical manifestation of harm. The actor need not even be aware of the physical manifestation. Perhaps the purpose is that it aids the fictive ideal observer of the strongly objective theory. Yet the ideal observer focuses exclusively on whether

one actor will, in fact, kill a second actor and if so, in fact, when. The focus is not on whether there is a physical manifestation indicative of an actor being about to kill, but whether an actor is, in fact, about to kill. The requirement is neither an infallible indicator of whether nor when. An actor may exhibit a physical manifestation a split second before killing or hours before killing. Moreover, an actor may exhibit a physical manifestation suggesting a forthcoming attack and nonetheless not attack. Even if it is argued that the requirement is correct in most cases, the strongly objective theory does not deal in likelihoods or probabilities or well-founded reasons for belief but in certitude. Moreover, why would the fictive ideal observer utilize an unreliable guide much less need a guide at all? Under a strongly objective theory either it is the case that an actor's right to life is about to be violated or not. What difference does it make whether the threat of harm is manifested physically? Moreover, what difference does it make when there is no requirement that the actor actually perceive it?

The PMR is an evidentiary device added *ad hoc* to an approach which is concerned with ontologically if and when an actor's right to life is, in fact, about to be violated. The requirement of an externalized threat of harm is used as a proxy for whether there is, in fact, a threat of harm. Incorporating that criterion into a strongly objective theory is elevating the proxy over the underlying principle, elevating epistemic justification over ontological truth. Since the proxy can both be present when the underlying principle is not and not present when the underlying principle is present, the proxy imperfectly translates the underlying principle. If and when an actor is about to kill, in fact, and if and when self-defense force is necessary, in fact, is not invariably a function of when an externalized threat of harm is present.

Yet without that *ad hoc*, unprincipled requirement, the strongly objective theory yields the absurd consequence of justifying villainous aggressors and forfeiting the lives of innocents. It yields conclusions about a paradigm case of

permissible self-defense diametrically opposite to our nearly undisputed intuitions. Although incorporating the ad hoc, unprincipled requirement avoids the absurd consequences, it nonetheless subverts and distorts the strongly objective theory of self-defense. It renders the strongly objective theory no longer strongly objective. Yet an internally consistent (i.e., without the PMR) strongly objective theory cannot accommodate our intuitions, even in paradigm cases.

Even if the inclusion of the PMR did not succumb to the difficulties demonstrated above and successfully avoided the counterintuitive outcome of justifying VA's aggression, some argument would need to be advanced as to why the PMR is other than an *ad hoc*, unprincipled requirement. If it was successful, clearly it would be useful in avoiding the *reductio ad absurdum* of yielding results diametrically opposite to our nearly undisputed intuitions in paradigm cases. But is there an independent rationale for the requirement? George Fletcher argues that a "visible manifestation of aggression"[34] by the aggressor or threat is necessary as it "signals to the community that the defensive response is not a form of aggression but a legitimate response in the name of self-protection."[35] Given Fletcher's theory's requirement that the actor know of the threat,[36] a physical manifestation makes it more likely that at least the actor claiming self-defense *could* have been aware of the threat. As a "signal," the PMR serves an evidentiary function: force could *be* permissible self-defense, but without such a signal we would not *know* that its permissible self-defense. As a result, force which objectively *is* permissible self-defense but is not *known* to be permissible self-defense is, under the PMR, impermissible self-defense. Regardless of whether one ultimately finds Fletcher's view persuasive, it is at least understandable that in a theory which requires that the actor know of the threat that there be something of which

34. George Fletcher, *Domination in the Theory of Justification and Excuse*, 57 U. PITT. L. REV. 553-78 (1996).

35. *Id.* at 571.

36. *See infra* note 49, authorities cited therein, and accompanying text.

the actor may have knowledge.[37] Fletcher further stresses that the issue "properly falls into the domain of political theory rather than moral theory."[38] But in Thomson's *moral* theory of self-defense which disregards the mental states of actors, a principled basis for the requirement seems to be lacking.

In addition to being inconsistent with a strongly objective approach and unprincipled, the PMR is both underinclusive and overinclusive. To see how the PMR is overinclusive, let us suppose that one actor exhibits an actual (not mistakenly perceived) physical manifestation that he is about to kill a second actor (who has a right to life) thus forfeiting the first actor's life and rendering force used against the first actor permissible. But, in fact, the first actor would have changed his mind at the last second and not have gone through with the attack that his physical manifestation suggested. Since he never would have killed the second actor even if not stopped, he never would have violated the other actor's right to life. But under the PMR, for example, A's physical manifestation of harm forfeits A's right to life rendering B's force against A permissible. Amending Thomson's theory to include the PMR would justify self-defense in situations in which the defender's right to life would never have been violated (even if the apparent attacker was not stopped). Actors who never would have violated another's right to life (even if not stopped) could nonetheless be permissibly killed. Such a result is antithetical to a strongly objective approach. Thomson

37. Not surprisingly, the two leading legal theorists who contend that self-defense may be permissible or justified without regard to subjective factors or mental states do *not* require a physical manifestation of harm in order for self-defense to be permissible or necessary. *See* PAUL ROBINSON, 2 CRIMINAL LAW DEFENSES 73-74, 76-79 (1984); GLANVILLE WILLIAMS, TEXTBOOK OF CRIMINAL LAW 504 (2d ed. 1983). For a critique of their strongly objective theories of legal justification see Russell Christopher, *Unknowing Justification and the Logical Necessity of the* Dadson *Principle in Self-Defense*, 15 OXFORD J. LEGAL STUD. 229 (1995).

38. Fletcher, *supra* note 34, at 570. For Thomson's acknowledgement of the moral theory/political theory distinction, in a different context, see *Imposing Risks*, *supra* note 3, at 176.

might properly respond to the overinclusiveness argument by revising the PMR to the following: a physical manifestation that would, in fact, have led to the death of another if not stopped.[39]

Yet even after incorporating the revised requirement (which defeats the overinclusiveness argument), Thomson's amended theory is nonetheless underinclusive. Suppose an evil aggressor, A, is a split second from killing innocent B yet has not exhibited a physical manifestation of being about to kill B. Under either version of the PMR, A is not yet about to violate B's right to life and thus has not yet forfeited his own right to life. Until A forfeits his right to life, B may not permissibly kill A. Yet suppose that by the time A does exhibit the requisite physical manifestation rendering it permissible for B to kill A in self-defense, it is too late for B to defend himself. B's life is in jeopardy from A, A is (at least temporally) about to kill B and defensive force is, in fact, immediately necessary for B to survive, but under either PMR force used at that point would be impermissible. This would seem to be the sort of case that Thomson's strongly objective theory would wish to permit B to save himself and kill A.

The incorporation of either PMR also produces arbitrary results. Consider the following hypothetical in which there are two villainous aggressors. Suppose at T-5, A forms a firm intention to murder B, who is driving down the road in A's general direction, with his antitank gun. Because A can kill B with only four seconds of preparation, he decides to wait until B gets closer so that he won't miss. At T0, ignorant of A's intentions, B forms a firm intention to kill A as soon as possible. But because he does not have his antitank gun, all he can do is try to run A over with his truck. B swerves into the other lane and starts driving at A. At T1, A readies his antitank gun and at T5 A fires the gun thereby killing B.

---

39. I am indebted to Kent Greenawalt for supplying this point.

Under Thomson's theory, amended to incorporate either PMR, B is the first to be about to violate A's right to life at T0 by driving at A. A does not exhibit a physical manifestation of being about to kill B until T1. Thus A's force is permissible in self-defense and B's force is impermissible. A's force is permissible, under Thomson's theory with either PMR, only because A exhibits a physical manifestation to be about to kill B subsequent to B's physical manifestation despite A being the first to intend to kill the other. In this hypothetical the permissibility of self-defense is a function of which party has the speedier method of killing. If A's method of killing B took longer to deploy and A exhibited a physical manifestation to be about to kill B at T-1 instead of T1, then it would be B's force that was permissible and A's impermissible.

A theory which produces arbitrary results (justifying the self-defense force of whichever actor has the quicker means of killing at his disposal) is especially prone to manipulation by knowledgeable criminals. One could plan to murder anyone and have her conduct endorsed as permissible self-defense as long as she made sure she killed her victim in such situations in which her means of killing could be more speedily employed than her victim. For example, suppose that B has an antitank gun after all. B has the top of the line model which only requires one second of preparation. Instead of killing by driving at A, B can wait until after T1 when A exhibits a physical manifestation (thereby forfeiting A's life), ready her gun at T2 and kill A at T3. Since A's physical manifestation at T1 occurred prior to B's at T2, A forfeits his right to life at T1 rendering B killing A permissible self-defense.

Access to a speedier method of killing hardly seems to be a principled basis for a theory to determine the permissibility or impermissibility of self-defense force. The incorporation of either PMR into Thomson's theory yields arbitrary results. Whether A or B is entitled to be justified in self-defense should not depend on which actor has the quicker

means of killing at their disposal. Why would a theory of self-defense favor the actor who can kill more quickly?

In addition to yielding arbitrary results, being subject to manipulation and, being underinclusive of the sorts of instances that Thomson would wish to permit the defender to kill the aggressor, the amendment is incompatible with the strongly objective account which Thomson employs. The amendment really serves as an evidentiary proxy best suited for a subjective theory.

VI.

Thus far we have been assuming that a PMR or amended PMR successfully avoids the counterintuitive result of justifying VA's force and forfeiting the right to life of innocent Vic. The subsequent issue has been whether the PMR raises other problems. Yet incorporating the PMR or amended PMR into the strongly objective theory yields a paradox that may prevent it from successfully avoiding the counterintuitive results.

Let us return to Thomson's case of Villainous Aggressor. Under the PMR or amended PMR, the first actor (between VA and Vic) to exhibit a physical manifestation is about to violate the right to life of the other and that first actor is subject to permissible force in self-defense from the other actor since that first actor has forfeited his own right to life. Since it seems that VA is the first to exhibit a physical manifestation, Vic's force is justified and it seems that inclusion of a PMR yields the intuitively expected outcome.

Yet if we assume that VA, under Thomson's theory with a PMR, is the first to exhibit a physical manifestation to be about to kill the other and is about to violate Vic's right to life rendering Vic's force permissible in self-defense, that premise paradoxically generates a conclusion contradicting the premise. To see the paradox, let us designate the point in time at which VA forfeits his right to life (by exhibiting a physical manifestation) and thereby force in self-defense against VA is permissible as T0. Also, let us assume that Vic cannot apply force against VA instantaneously (i.e.,

without any time elapsing) and that all applications of force necessarily entail a prior physical manifestation.[40] Thus in order for Vic to do that which he may permissibly do (apply lethal force to VA at T0),[41] he would necessarily have had to exhibit a physical manifestation at some point in time prior to his permissible application of force in self-defense at T0. Since Vic exhibits a physical manifestation prior to T0 and VA's physical manifestation occurs at T0, it is Vic who is the first to exhibit a physical manifestation of being about to kill the other and Vic who is about to violate the right to life of VA. Thus it is Vic who forfeits his right to life rendering VA's force permissible in self-defense and Vic's force impermissible.

Perhaps this claim warrants further explanation. Since our premise is that VA has lost his right to life at T0, it will not violate VA's right to life to be killed (or sustain lethal force) at T0. For example, if a bullet from Vic's gun enters VA's skull at T0 (or a fraction of a nanosecond after T0)[42] that should be permissible since it does not violate VA's right to life since VA has lost his right to life. Or, for example, if Vic squeezes the trigger of his gun at T0 which causes a bullet to pierce VA's skull and kill him that should also be permissible for the same reason. In either case, Vic would have necessarily exhibited a physical manifestation (taking the gun out, raising it, cocking it, aiming it etc.) of his permissible force prior to T0. For example, in order for Vic to pull the trigger at T0 he would have had to do other acts like taking the gun out, raising it etc., any of which might constitute a physical manifestation, prior to T0.

---

40. If this is not the case then support for a PMR erodes–why would there be the requirement in the first place?

41. Although Vic is not required to apply force at T0–he could do it at T5 as in the previous discussion or T3, T7 etc.–nonetheless it is permissible for him to do it at T0.

42. It is not necessary, in order for the paradox to arise, that VA's exhibition of a physical manifestation and Vic's application of force occur simultaneously. Vic's application of force could occur subsequent to T0 as long as it was a sufficiently short period of time thereafter such that the duration of any physical movement of Vic's that would satisfy a PMR was of longer duration.

The premise that Vic's force in self-defense is permissible, with which Thomson and nearly everyone would intuitively agree, yields the conclusion that Vic's force is impermissible and VA's conduct permissible. Thus in doing what we assume Vic may permissibly do, under either PMR, Vic exhibits a physical manifestation prior to VA's physical manifestation at T0. As a result, it is Vic who is the first to exhibit a physical manifestation and thus Vic who is about to violate the right to life of VA and it is Vic who forfeits his own right to life. It is VA's conduct which is permissible and Vic's force that is impermissible. In other words, by incorporating a PMR or amended PMR into Thomson's strongly objective theory, in doing what is permissible an actor paradoxically does what is impermissible. Inclusion of the PMR or amended PMR not only results in the same counterintuitive outcome as the strongly objective theory without a PMR, but it also produces a paradox: in order to do what is permissible that which is permissible becomes impermissible.

Of course, Thomson could avoid the paradox by scrapping the PMR or amended PMR. Yet this would result in the counterintuitive outcome of forfeiting Vic's right to life and justifying VA's conduct, as discussed in section IV. The PMR or amended PMR need not be abandoned, however, to avoid the paradox. Thomson's theory could adopt an additional amendment. A waiting period after the aggressor or threat has lost his right to life could be imposed before the defender could use force. The waiting period would have to be of sufficient duration such that the defender could not exhibit a physical manifestation of his defensive force until after the aggressor's or threat's physical manifestation. In other words, even though force in self-defense is permissible as soon as the aggressor or threat forfeits his right to life, the defender must wait until some amount of time has passed before using force in self-defense.

Although this waiting period amendment technically circumvents the paradox, it is *ad hoc*, unprincipled and lacking an independent rationale. If force in self-defense is

permissible as soon as the aggressor or threat loses his right to life, why would the defender need to wait? If the aggressor or threat has already lost her right to life what morally recognized interest of the aggressor is being served by waiting? Is the aggressor's no longer existent right to life being violated less by waiting?

The awkwardness of incorporating this amendment into an objective theory should be contrasted with the conceptual ease in which (at least partially) agent-centered subjective approaches naturally incorporate what functions as a waiting period—reasonable belief or knowledge of the justificatory circumstances. Under subjective approaches which require some requisite mental state of the defender claiming self-defense, the above paradox never arises because satisfying the requisite mental state acts as a functional substitute for the ad hoc waiting period. Consider the following example. A exhibits a physical manifestation of being about to kill B. Unlike under Thomson's theory with a PMR, force by B in self-defense is still not yet permissible under an at least partially agent-centered approach. B must form, for example, an honest or reasonable belief regarding, or have knowledge of, A's physical manifestation before B's force can be considered permissible. The time it takes B to perceive and process B's physical manifestation before responding insures that B's physical manifestation will not occur prior to A's. Therefore, it will not be the case under such subjective approaches, that in doing what is permissible, the defender will do what is impermissible.

I believe that the latent paradox has been overlooked because our shared conception of self-defense depends on an at least partially subjective account. Perhaps even proponents of an objective approach unwittingly or subconsciously slide into or assume a subjective view which has the effect of obscuring the paradox. Only by fully explicating the ramifications of an objective approach does the paradox come to light. So ingrained in our intuitive understanding of self-defense is a subjective view that we naturally assume what amounts to as a waiting period even in

an objective approach which does not make allowance for one.

Central to our shared understanding of self-defense, I believe, is that the defender's force is a *response* to the threatened harm of the aggressor or threat. In order for the defender to respond, however, entails the defender having some mental state of perception, belief, knowledge etc. in regard to the threat posed. In treating such mental states as irrelevant, an objective account obscures an essential aspect of self-defense—that it be a response. Whereas an at least partially agent-centered account affirmatively requires some mental state which insures that the defender is responding, an objective approach treats the defender's force *as if* it was a response. The paradox discussed above arises in the crack between requiring that the defender actually respond and treating the defender as if she responded.

VII.

Perhaps Thomson intended to employ not a strongly objective approach but rather a weakly objective[43] approach. Under the latter approach, in cases in which the threat or aggression is uncertain, conduct is impermissible if there is a sufficiently high probability that the aggression or threat posed will violate the right to life of another. Force in self-defense is permissible against such sufficiently high probability, though not certain, aggression or threats. Would the PMR incorporated into the weakly objective approach avoid the difficulties caused by the inclusion of the requirement in the strongly objective approach? The problems of arbitrariness, the paradox, underinclusiveness and overinclusiveness would still persist. Though undesirable, under and overinclusiveness would not be as grave a problem for an account that determines self-defense to be permissible not in all cases in which, in fact, an actor's right to life is

43. See the discussion of the distinction between strongly and weakly objective approaches, *supra* section II.

about to be violated but only when it was of sufficiently high probability. Under and overinclusiveness is inevitable in such a probabilistic account.

Additional problems, however, remain. Is the physical manifestation requirement an absolute indicator of sufficiently high probability or is it merely one factor that goes into the calculus of sufficiently high probability? If its only one factor among many, then in a situation in which the physical manifestation is not present but there was nonetheless a sufficiently high probability, then the weakly objective approach would have the same problem that the strongly objective theory without the PMR incurred. It would yield results in paradigm cases inapposite of our virtually undisputed intuitions.[44] On the other hand, if the PMR is an absolute indicator, then in a situation in which the requirement was satisfied but there was nonetheless not a sufficiently high probability the requirement would contradict the essence of the very approach (high probability) of which it is but a part.

Perhaps incorporating the amended PMR (physical manifestation that would, in fact, lead to the death of another unless stopped) into the weakly objective theory would surmount the above difficulties. Yet the requirement of a physical manifestation leading to *certain* death unless stopped is incompatible with an approach that involves only *uncertain*, but high probability, violations of rights to life.

The most serious objection to the claim that Thomson might have intended to employ only a weakly objective theory is that such a theory is only applicable to cases of uncertain harm or uncertain violations of rights to life. All of the cases being discussed herein involve only *certain* harms or violations. Moreover, Thomson herself only discusses instances of certain threats and certain aggression in *Self-Defense*. She explicitly points out that the problem of uncertain threats and aggression is outside the scope of her article.[45]

---

44. *See supra* section IV.

45. *See supra* note 22 and accompanying text.

Perhaps Thomson intended neither a strongly objective nor weakly objective theory but rather a minimally objective account. The minimal condition for a theory to be objective, I take it, is that it exclude subjective criteria. Michael Gorr has provided the helpful term of "purely externalist"[46] for an account that excludes all internal aspects of actors, i.e., mental states, but focuses on the actors' external conduct. The advantage to Thomson of this approach is that it could avoid the interpretation of about to as strictly temporal. Unlike under a strongly objective theory, being temporally the first to be about to kill would not necessarily constitute being about to violate another's right to life. Only some external, physical act or conduct of an actor could make that actor be about to violate another's right to life. Such a "purely externalist" account would feature some type of PMR. Also unlike the strongly objective approach, the inclusion of a PMR would not render the purely externalist account internally inconsistent.

Yet the other problems of the PMR, discussed above, including the paradox, would still remain. The minimally objective or "purely externalist" account fares no better than the weakly objective theory. The next section will consider a number of variations of subjective theories and will assess whether they avoid the counterintuitive result of VA permissibly killing Vic discussed in section IV.

VIII.

Although there are any number of subjective theories that could be considered, we will consider three. The first, a target-centered account, takes into account subjective aspects of the actor who has conventionally been designated the aggressor or threat, e.g., VA or IT. The second two, at least partially agent-centered accounts, focus on subjective characteristics of who we typically consider the defender.

Larry Alexander has criticized Thomson's theory, in part, for its justifying force against moral, but causally

46. Michael Gorr, *Private Defense*, 9 LAW & PHIL. 254 n.28 (1990).

harmful, innocents.[47] He questions why we should favor one moral innocent, e.g., Vic over another, e.g., IT. Alexander rejects Thomson's principle that every person has the right not to be killed unless they are the first to be about to violate another's right to life and instead proposes that everyone has a right not to be killed culpably.[48] Although it is unclear whether Alexander's theory also incorporates the latter portion of Thomson's principle, for the purposes of ascertaining where Thomson's theory goes awry let us assume that under Alexander's theory one's right is also qualified by "unless they are the first to be about to violate another person's right not to be killed culpably."

Alexander believes that one's right not to be killed culpably is violated when the killer is aware or believes that his conduct is wrongful. At the risk of oversimplification, Alexander's right may reduce to that one has a right not to be killed by non-innocent (evil or culpable) threats and aggressors. Thus one lacks a right not to be killed by moral, though causally harmful, innocents. For example, IT, falling toward Vic, is not culpably about to violate Vic's right to life. Self-defense force by Vic against IT is impermissible under Alexander's theory.

Let us now consider whether Vic may use self-defense force against VA. At T0, VA is about to culpably violate Vic's right to life unless Vic was already about to culpably violate VA's life at T-5. Although Vic, at T-5 may have been about to violate VA's right to life, he was not culpably about to violate VA's right to life. Thus at T0, VA is the first to be about to culpably violate the other's (Vic's) right to life and VA thereby forfeits his right to life. Since Vic will otherwise be killed by VA, and Vic is not about to culpably violate VA's right to life because VA has already forfeited it (and Vic is not culpable), self-defense force by Vic is permissible.

47. Alexander, *supra* note 2, at 60-62.
48. *Id.* at 60.

Like Thomson's and Alexander's theories, Fletcher's theory[49] requires that an actor, in order for her force to be permissible in self-defense, face an objectively actual threat. A mistaken belief, even if reasonable, that aggression is imminent does not suffice. Yet unlike Thomson's strongly objective theory, Fletcher adds the (subjective) requirement that the actor have knowledge of the justificatory circumstances or act with justificatory purpose. Does Fletcher's theory avoid the reductio ad absurdum of justifying VA's force against Vic? Since VA acts with villainous intent rather than with justificatory purpose, VA's force would be unjustified; since Vic correctly perceives VA's threat and acts with justificatory purpose, Vic's force would be justified. The inclusion of a subjective component in Fletcher's theory prevents the (presumably) absurd result of justifying the villainous aggressor against the innocent defender.

A third (at least partially) subjective theory requires, in order for an actor to be justified in self-defense, that she reasonably believe that her force is necessary against an unjustified threat.[50] Would Vic be justified against VA under Thomson's theory with this reasonable belief amendment? Seeing VA driving at him to run him over and kill him, Vic might believe that force was necessary in self-defense. Vic might also believe that VA's threat is unjustified since Vic realizes that he has done nothing to provoke

---

49. Fletcher's theory of justification can be found in, among other works, GEORGE FLETCHER, RETHINKING CRIMINAL LAW 552-79, 759-875 (1978); George Fletcher, *The Nature of Justification, in* ACTION AND VALUE IN CRIMINAL LAW 175 (Stephen Shute et al. eds., 1993); *The Right Deed for the Wrong Reason*, 23 UCLA L. REV. 293 (1975).

50. This approach to justified self-defense is embodied in the influential Model Penal Code of the American Law Institute. MODEL PENAL CODE AND COMMENTARIES § 3.04 (Official Draft and Revised Comments 1985). Support for this account can be found in Russell Christopher, *Mistake of Fact in the Objective Theory of Justification: Do Two Rights Make Two Wrongs Make Two Rights...?*, 85 J. CRIM. L. & CRIMINOLOGY 295 (1994); Joshua Dressler, *New Thoughts About the Concept of Justification in Criminal Law: A Critique of Fletcher's Thinking and Rethinking*, 32 UCLA L. REV. 61 (1984); Kent Greenawalt, *The Perplexing Borders of Justification and Excuse*, 84 COLUM. L. REV. 1897 (1984).

VA and it is not he but rather VA who is at fault in initiating the confrontation. If Vic's beliefs are reasonable, he would be justified in self-defense. Even if we supposed that VA honestly believed that driving at Vic to run him over and kill him was necessary in self-defense and that Vic constituted an unjustifiable threat, VA would lack reasonable grounds for his belief. Like Alexander's and Fletcher's theories, the Model Penal Code's account of justifiable self-defense avoids the counterintuitive outcome of rendering Vic's force impermissible and VA's force permissible.

Under neither of the three above (at least partially) subjective theories would VA's aggression be permissible and Vic's force impermissible. This suggests that the feature of Thomson's theory producing the unfortunate result of a villainous aggressor being justified against a moral innocent is the lack of consideration of subjective factors.

If the three (at least partially) subjective theories incorporated either version of the PMR only the (at least partially) agent-centered subjective theories would avoid the paradox discussed in section VI. Target-centered accounts, which do not require any particular mental state of the defender asserting a self-defense justification, incorporating some variant of a PMR, would incur the paradox.[51]

CONCLUSION

Although we may all agree (on an intuitive level) in a paradigm case such as that of Villainous Aggressor that VA's conduct is impermissible and Vic's conduct is permissible self-defense, the theoretical challenge is to devise a theory which yields the same conclusion as our nearly undisputed intuitions. The essence of permissible self-defense is that it is a response to a previous impermissible threat of harm. Yet in order for the force employed in self-defense to successfully prevent the threatened harm from actually occurring it must, at least in some cases, be employed prior to

51. It is unclear whether Alexander's target-centered theory includes some type of PMR.

the fruition of the threatened harm. The difficulty lies in identifying which force is defensive and which is aggressive in such cases when what we intuit to be the defensive force is employed first. Any viable theory of self-defense must be able to account for why the initial application of force is not impermissible aggression but rather permissible self-defense.

Thomson's theory of self-defense, embedded in a strongly objective account, allows the condition for identifying the impermissible aggressor (the first to be about to violate the right to life of another) to be interpreted as strictly temporal. Since who we intuit to be the innocent defender is, in some cases, temporally closer to killing than the aggressor or threat, the innocent defender is the first to be about to violate the right to life of the wrongful aggressor. Contrary to our intuitions and Thomson's conclusions, the innocent defender's conduct is deemed impermissible and the wrongful aggressor's force impermissible. To avoid the strictly temporal interpretation of "about to" resulting in a reversal of our intuitions, about to must be interpreted as (or Thomson's theory amended to include) some type of PMR. Yet inclusion of a PMR is inconsistent with the strongly objective approach; the problem of underinclusiveness and the evidentiary nature of a PMR renders the strongly objective account no longer strongly objective. Inclusion of a PMR would be less problematic, however, with either a weakly objective or "purely externalist" theory, but any type of PMR in an objective account results in the paradox in which what is permissible is impermissible. The paradox may be technically circumvented by a requirement of a waiting period, but the requirement lacks any satisfactory independent rationale. Unless a principled basis for the waiting period amendment can be supplied, Thomson's theory can only provide a satisfactory account of self-defense by including subjective criteria.

A subjective account assesses some mental state of one or more of the combatants. In determining which party was the wrongful aggressor and which the innocent defender, it

inquires into the beliefs, intentions or states of knowledge of the actors. In the case of Villainous Aggressor, Vic can point to his (reasonable) belief that he perceived that VA was about to kill him and he had no intention to kill VA until after he perceived VA's attack. VA neither believed nor could he reasonably believe that Vic intended to harm him until after he commenced his attack on Vic. Despite Vic being the one to use force first and being (at least temporally) about to use force first, a subjective theory can easily distinguish between the conduct of the aggressor and the innocent defender by examining the beliefs,intentions or other mental states of the actors. A strongly objective theory, on the other hand, fails to successfully distinguish threats/aggressors from innocent defenders.

Any objective account must distinguish the impermissible threat or aggression from the permissible self-defense to which it is in response without recourse to subjective factors such as belief, intention or knowledge. It must identify some asymmetry or some type of act, conduct or state of affairs regarding who we intuit to be the aggressor as constituting the initial or first impermissibility.

Other than subjective factors, the possible candidates are unsatisfactory. The first application of force is not viable because who we intuit to be the innocent defender may well be the first to apply force. The first actor to be about to use force, if interpreted strictly temporally, is necessarily the first actor to actually use force. Since the first actor to actually use force may well be who we intuit to be the innocent defender, the state of affairs of being the first to be about to use force also fails. The first physical manifestation of harm or the amended PMR, in addition to a host of other problems, yields the paradox of if the defender's conduct is permissible then its impermissible. What other candidates remain for identifying some conduct or aspect of who we intuit to be the aggressor or threat in paradigm cases as the first or initial impermissibility? What asymmetry (with a principled rationale) allows us to distinguish

between the impermissible threat or aggressor and the permissible self-defense of the defender under an objective account?

It would seem that our traditional conception of self-defense can only be accounted for by resort to one or more subjective criteria. An objective account of self-defense appears to be suspect.[52]

52. Since canvassing every extant formulation of an objective theory of self-defense is outside the scope of this article, I have not demonstrated that all objective accounts are problematic. I would merely like to suggest that in light of the difficulties raised it is worth entertaining the possibility that any principled objective account will fail to yield results compatible with our intuitions in paradigm cases.