The Folk Ethics of Self-Defense: An Emprical Study on the Moral Permissibility of Killing Apparent Threats

August Renbo Olsen (August.Olsen@fil.lu.se)
Lund University. Department of Philosophy, Helgonavägen 3, 22222 Lund, Sweden

Jan García Olier (jangarcia.olier@outlook.com)
University of Zurich, Department of Philosophy, Zollikerstrasse 117, 8008 Zurich, Switzerland

Pascale Willemsen (pascale.willemsen@uzh.ch)
University of Zurich, Department of Philosophy, Zollikerstrasse 117, 8008 Zurich, Switzerland

Abstract

Philosophers of self-defense debate whether it can be morally permissible to kill an aggressor who only *appears* to threaten you. In developing moral theories of self-defense, these philosophers sometimes make (untested) conjectures about what most people believe about self-defense. This paper aims to explore lay judgments on this issue. To do so, we conduct three pre-registered experiments manipulating the actuality of a threat. Across abstract and concrete scenarios as well as within-subjects and between-subjects designs, results consistently show that laypeople judge certain self-defensive killings morally permissible regardless of whether the aggressor poses a genuine threat or a merely apparent threat. These findings oppose Objectivist views on self-defense, which hold that self-defense is only permissible when facing a genuine threat. Instead, they support Subjectivism and what we call the "It's Complicated View", both of which hold that apparent threats can justify lethal self-defense (albeit with possible variation in permissibility ratings).

Keywords: ethics of self-defense; subjectivism; objectivism; ethics of war; moral psychology; experimental philosophy.

Introduction

Self-defense often involves harming or even killing someone to prevent them from harming you. Using lethal force in defense against threats to one's life, grave bodily injury, kidnapping, or rape, is a standard criminal law defense in the US (see, for instance, Section 3.04 of the US Model Penal Code) and countless other countries (e.g., Germany, England, Japan, Spain, and Colombia).

However, as harming and killing are serious matters, this raises questions about the conditions under which self-defense is morally permissible. Most philosophers working on the ethics of self-defense agree that it may be morally permissible to kill someone in self-defense if the person would otherwise kill you (see, e.g., Doggett, 2011; Coons & Weber, 2016; Frowe & Parry, 2022). However, philosophers disagree on whether it is morally permissible to kill an aggressor who merely *appears* to threaten one's life (see, e.g., Ferzan, 2005; McMahan, 2005; Frowe, 2010). For example, suppose Aggressor points a gun at Defender and makes it clear that he is about to pull the trigger. Unbeknownst to Defender, Aggressor only holds an unloaded gun and cannot kill Defender. Defender believes (on reasonable grounds) that their life is in danger. The question philosophers disagree on, then, is: *given that Aggressor merely appears to threaten Defender's life, is it morally permissible for Defender to kill Aggressor?*

Within the debate, philosophers discuss how well their own theory of self-defense accords with 'common opinion.' All else being equal, it is often taken as counting in favor of a theory if it concurs with people's moral intuitions about self-defense. By appealing to this common opinion, philosophers make empirically testable conjectures about 'what most people believe.' Unfortunately, these philosophers sometimes disagree on 'what most people believe', and they do not provide any empirical evidence supporting their hypotheses.

This paper aims to remedy this by testing how laypeople evaluate the permissibility of killing apparent threats. We conduct three pre-registered experiments, manipulating whether the aggressor poses a *genuine* threat to the defender's life or merely *appears* to pose a threat. Our results clearly support the view that the actuality of a threat, though potentially morally relevant, does not affect whether a lethal reaction on the victim's part is judged permissible.

Theoretical Background

We distinguish two leading positions in the ethics of self-defense: *Objectivism* and *Subjectivism*.[1] Both positions accept what we may call the *Threat Condition:* it is only permissible to harm someone in self-defense if someone *poses a threat* to the defending agent. Self-defense is, per

[1] As we will use these terms, Objectivism has been defended by, among others, Judith Jarvis Thomson (1991) and Jeff McMahan (2005) (though see McMahan (2011) for remarks suggesting Subjectivism). It has been speculated that many philosophers of self-defense implicitly 'slant' towards Objectivism (Frowe, 2010). Subjectivism has been defended, among others, by Kimberly Ferzan (2005), Helen Frowe (2010) and Renée Jorgensen Bolinger (2021).

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

definition, a defense against some threat.[2] However, Objectivism and Subjectivism differ in what counts as 'posing a threat'.[3] This disagreement leads them to treat cases where the threat is only apparent differently. Recall our previous example where Aggressor is about to pull the trigger of a gun pointed at Defender, the gun is unloaded, but Defender believes it is loaded. Is Aggressor posing a threat to Defender's life here? This depends on what notion of 'threat' is applied. Aggressor is not posing a 'genuine threat' to Defender's life, as Aggressor's gun is unloaded and Defender's life is therefore not *actually* in danger. However, Aggressor is posing an 'apparent threat' to Defender's life, as it *appears* to Defender as if Aggressor is about to kill him.

Objectivism only permits self-defense against genuine threats. That is, for self-defense to be permissible, Defender has to *actually* be in danger (see, e.g., McMahan, 2005). Thus, it is impermissible for Defender to kill Aggressor when aggressor only poses an apparent threat, even if Defender had reason to believe that the threat was genuine, but, through no fault of his own, this belief was false. *Subjectivism*, on the other hand, agrees that it is sometimes permissible to kill genuine threats, but also holds that it is sometimes permissible to kill merely apparent threats. Both Objectivism and Subjectivism further hold that Defender would normally not be blameworthy for killing an apparent threat in the circumstances we describe, where Defender reasonably believes Aggressor poses a real threat.[4]

Philosophers of self-defense are interested in answering a normative ethical question, namely, under what conditions killing in self-defense is morally permissible. We do not aim at answering this ethical question using experimental methods. We merely aim at uncovering laypeople's intuitions about the permissibility of killing apparent threats. As mentioned earlier, however, some philosophers of self-defense think it counts in favor of a moral theory if it aligns with common opinion, and thus, our findings might be used to motivate one moral theory of self-defense over another.

What, then, would it take for common opinion to align with Objectivism or Subjectivism? If laypeople were asked to evaluate cases of self-defense using a scale reaching from 'clearly impermissible' to 'clearly permissible', Objectivism holds that killing merely apparent threats in self-defense should receive a mean rating somewhere on the 'impermissible' side of the scale, whereas self-defense against genuine threats should be on the 'permissible' side. In contrast, Subjectivism holds that both killings against genuine and apparent threats should receive a mean rating somewhere on the 'permissible' side.

While the dominant philosophical accounts fall into Objectivism or Subjectivism, we want to construe a third view, which we call the *It's Complicated View* (ICV). ICV does not conceive of 'permissibility' as a binary concept, but allows that an action can be *more or less* permissible than another action. ICV holds that both killing genuine threats and killing apparent threats in self-defense should be judged permissible, but that killing genuine threats should be judged *more permissible* than killing apparent threats. ICV thus sits comfortably in the middle between Objectivism and Subjectivism. It agrees with Subjectivism that killing apparent threats can be permissible, and it agrees with Objectivism that the genuineness of a threat marks a morally relevant difference.

Key Hypotheses

With this theoretical background in mind, we conducted three pre-registered experimental studies. For the purposes of this paper, we will limit the presentation and analysis of our experiments to four main hypotheses:

(H1): Objectivism, Subjectivism, and the ICV alike predict that in the genuine threat condition, mean permissibility ratings are above the neutral midpoint of 5.

(H2): Subjectivism predicts that in the apparent threat condition, mean permissibility ratings are still above the neutral midpoint (5).

(H3): Objectivism predicts that in the apparent threat condition, mean permissibility ratings are below the neutral midpoint (5).

(H4): ICV predicts a significant main effect of Threat on permissibility ratings, such that permissibility ratings are lower in the apparent threat condition than in the genuine threat condition.

In all three studies, we recruited participants via Prolific Academics. All participants had a minimal approval rate of 90% and were English native speakers aged 18 or above, living in Australia, Canada, Ireland, New Zealand, or the United Kingdom. We excluded participants from the United States of America given their permissive legislation on civilians baring arms, and the public having strong and very polarized views on the right to defend oneself.

Study 1 a and 1 b

Philosophers of self-defense rely on philosophical thought experiments to support their theoretical arguments. These thought experiments vary significantly in how much detail they provide, ranging from largely abstract to very concrete descriptions of the situation (for discussion of concrete and abstract case designs and the different intuitions they elicit, see Nichols & Knobe, 2007). Both approaches have their

[2] Both positions accept another condition which we may call the *Responsibility Condition*: it is normally only permissible to harm aggressors who are *responsible* for posing a threat to the defender. We will not discuss what notion of responsibility is relevant here as philosophers disagree (for an overview, see Frowe & Parry, 2022). More importantly, in all our scenarios, the aggressor is responsible in each of the different senses philosophers find relevant.

[3] Notice that there is a difference between *posing* a threat (being about to harm someone) and *uttering* a threat (saying something threatening). Self-defense ethics focuses on *posing threats,* such as when one agent points a gun at another person and is on track to pull the trigger.

[4] For Objectivism, this is because Defender has an epistemic excuse.

merits but also the potential to pump unreliable intuitions, especially when used in *empirical* experiments with non-philosophers. While abstract cases are economical and can focus on the facts that philosophers consider relevant, it is often this oversimplicity that participants find challenging, and that can lead to uncontrollable additional assumptions and confounders. Scenarios rich in detail, on the other hand, may lead people's imagination astray and distract them from the information that actually matters to the question.

As neither approach is perfect, we test people's intuitions in response to both abstract and concrete stimuli and aim for converging evidence. To this end, we conducted two experimental surveys, presenting participants with either abstract (Study 1a) or concrete (Study 1b) versions of self-defense scenarios. In both studies, participants answered the following three questions presented in a fixed order on the same page:

1. **Moral permissibility**: From a moral point of view, to what degree do you think that Person B killing Person A was morally impermissible or permissible? (from "1 = clearly impermissible" to "9 = clearly permissible)[5]
2. **Blameworthiness**: How blameworthy is Person B for killing Person A? (from "1 = not blameworthy at all" to "9 = extremely blameworthy")
3. **Reasonability**: To what degree do you think it was reasonable for Person B to believe his life was in danger? (from "1 = not reasonable at all" to "9 = entirely reasonable")

Study 1a: Abstract Scenario

For this experiment, we recruited 202 participants via Prolific (gender-balanced; M_{age} = 40.1). We pre-registered the design and all hypotheses.

Materials and Procedures We implemented a 1 × 2 between-subjects design with Threat (genuine threat, apparent threat) as a between-subjects factor.

Results and Discussion. As pre-registered, we conducted a series of planned comparisons examining the impact of Threat on permissibility, blameworthiness, and reasonability ratings. One-sample t-tests revealed that permissibility ratings in both the Genuine and Apparent threat conditions were significantly *above* the neutral mid-point (all *ps* < *.001*, all *ds* > *1.28*), and so were reasonability ratings across conditions (all *ps* < *.001*, all *ds* > *2.73*). Blameworthiness ratings, on the other hand, were significantly *below* the neutral mid-point across Threat conditions (all *ps* < *.001*, all *ds* > *.64*). Independent samples t-tests further revealed a small yet significant effect of Threat on permissibility ratings (*p*= *.025*, *d*= *.27*) but no significant effect on blameworthiness (*p*= *.161*, *d*= *.14*). Table 1 shows the results in greater detail. Figure 1 depicts the mean permissibility ratings, the distributions, and the standard error of the mean across Threat conditions in Study 1a.

Table 1: Means and standard deviations in **Study 1a**

Threat	Permissibility		Blame		Reasonability	
	Mean	SD	Mean	SD	Mean	SD
Genuine	7.73	1.50	3.17	2.35	8.36	1.23
Apparent	7.28	1.77	3.50	2.33	8.66	.87

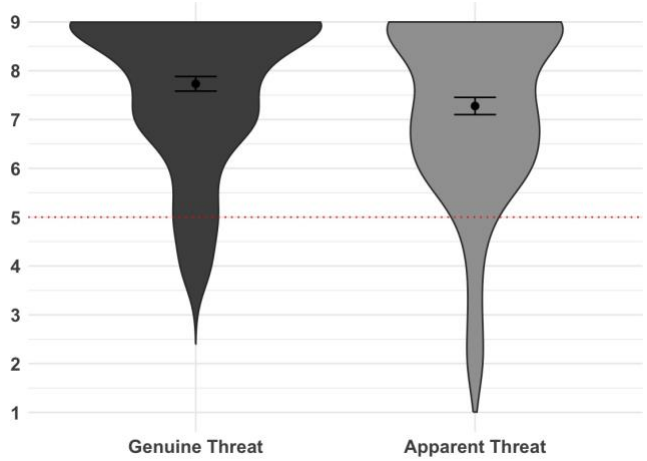

Figure 1. Violin plot for permissibility ratings in **Study 1a**. Error bars show the standard error of the means.

Study 1b: Concrete Scenarios

For this experiment, we recruited 331 participants (gender-balanced; M_{age} = 41.3). As in Study 1a, we pre-registered our hypotheses, described in greater detail in the pre-registration.

Materials and Procedures. We implemented a 3 × 2 full-factorial design with Story (Cleaner, Vet, Defibrillator) and Threat (genuine threat, apparent threat) as between-subjects factors. Participants were randomly assigned to one of the six conditions and, as in Study 1a, answered the three questions about permissibility, blameworthiness, and reasonability on the same page.

Results and Discussion. Following our pre-registration, we considered the Story a random factor and aggregated the data from the three Story conditions, treating the experiment as a 1 × 2 (Threat) design. In line with our hypotheses and the previous study, we conducted planned comparisons examining the impact of Threat on permissibility, blameworthiness, and reasonability ratings. Consistent with Study 1a, we found that permissibility ratings in both the Genuine and Apparent threat conditions were significantly *above* the neutral mid-point (all *ps* < *.001*, all *ds* > *1.33*), as were reasonability ratings (all *ps* < *.001*, all *ds* > *3.76*). Also consistent, we found that blameworthiness ratings across threat conditions were significantly below the neutral mid-

[5] The 9-point Likert scale participants were presented with for the moral permissibility question was anchored at -4 ("clearly impermissible") and 4 ("clearly permissible"). For comparability with the two other questions, however, the scale was re-coded from 1 ("clearly impermissible") to 9 ("clearly permissible").

point (all *ps* < *.001*, all *ds* > *.56*). Nevertheless, and in contrast to Study 1a, we observed no impact of Threat on permissibility ratings (*p* = *.128*, *d* = *.12*) or blameworthiness (*p* = *.128*, *d* = *.012*). Table 2 shows the results in greater detail. Figure 2 depicts the mean permissibility ratings, the distributions, and the standard error of the mean across Threat conditions in Study 1b.

The results provide greater support for Subjectivism. In both abstract and concrete self-defense scenarios, the fact that the threat was only apparent, not genuine, does not make laypeople judge the killing in self-defense as impermissible, nor the defender as blameworthy or unreasonable. However, consistent with ICV, we also found a small, yet significant, main effect on permissibility ratings when presenting participants with the abstract self-defense scenario.

Table 2: Means and standard deviations in **Study 1b**

Threat	Permissibility		Blame		Reasonability	
	Mean	SD	Mean	SD	Mean	SD
Genuine	7.64	1.51	3.91	2.60	8.54	.85
Apparent	7.43	1.82	3.59	2.50	8.58	.95

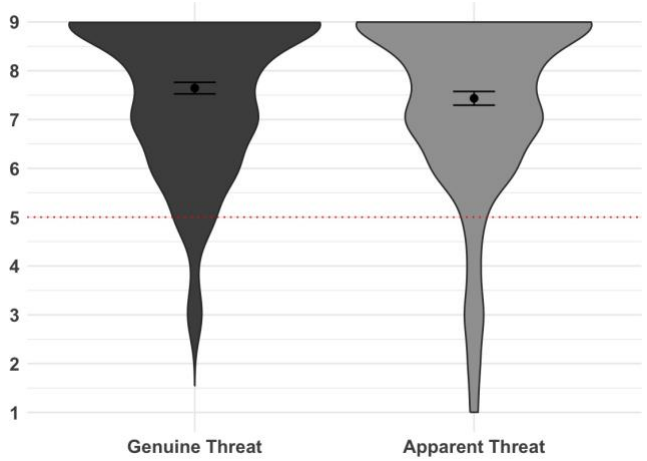

Figure 2. Violin plot for permissibility ratings in **Study 1b**. Error bars show the standard error of the mean.

Study 2

In our first studies, participants were only presented with a single scenario where the aggressor posed either a genuine or an apparent threat to the defender's life. Philosophers leaning towards Objectivism may object that this between-subjects experimental design does not adequately reflect the methodology of philosophical thought experiments. In philosophy, we do not give one version of a case to Philosopher A and another version to Philosopher B and then compare their answers. Rather, each philosopher would get both versions of the case, consulting their intuition while comparing the cases side by side. The objectivist might argue that if we had asked about the permissibility of killing genuine and apparent threats in self-defense in a within subjects-design, we would have found objectivist responses. Study 2 addresses this worry.

For this experiment, we recruited 350 participants (gender-balanced; M_{age} = 40.1).We pre-registered the design and all hypotheses.

Materials and Procedures. We implemented a 3 × 2 mixed design with Story (Cleaner, Vet, Defibrillator) as a between-subjects factor and Threat (genuine threat, apparent threat) as a within-subjects factor. The Threat conditions were presented randomly—either the genuine or the apparent threat first. Participants were randomly assigned to one of the six conditions, and after being presented with each of the within-subjects treatments, they were given the same three questions about moral permissibility, blameworthiness and reasonability as participants in study 1a and 1b.

Results and Discussion

We considered the Story a random factor and aggregated the data from the three Story conditions. In line with our pre-registered hypotheses, we examined the impact of Threat on *permissibility* ratings. Before moving on, however, we also analyzed whether there were any order effects. To this end, we first conducted a mixed ANOVA with Threat (genuine vs. apparent) as a within-subjects factor and Order (genuine first vs. apparent first). The ANOVA revealed no significant interaction effects between Threat and Order on permissibility ratings (*F(1, 348) = 2.033; p = .155;* η_p^2 *= .006*). The ANOVA also revealed a main effect of Threat *F(1, 348) = 83.101; p < .001;* η_p^2 *= .193*) but no effect of Order (*F(1, 348) = 1.248; p = .265;* η_p^2 *= .004*) on permissibility ratings. Considering we did not find any order effects, we proceeded as pre-registered and treated the experiment as a 1 × 2 (Threat) within-subjects design. We conducted a series of pairwise comparisons. Coherent with the ANOVA above, a paired samples t-test revealed a significant effect of Threat on permissibility ratings (*p < .001, d = .48*). When presented with both threat conditions, participants were more likely to say that killing genuine threats in self-defense was permissible (*M = 8.07*; *SD = 1.391*) compared to killing merely apparent threats (*M = 7.46*; *SD = 1.838*). Importantly, however, mean permissibility ratings were significantly above the neutral mid-point (5) regardless of whether the killing in self-defense was against a genuine threat (*p < .001, d = 2.20*) or a merely apparent one (*p < .001, d = 1.34*). Table 3 shows the results in greater detail. Figure 3 depicts the mean permissibility ratings, the distributions, and the standard error of the mean across Threat conditions in Study 2.

Table 3: Means and standard deviations in **Study 2**

Threat	Permissibility	
	Mean	SD
Genuine	8.07	1.391
Apparent	7.46	1.838

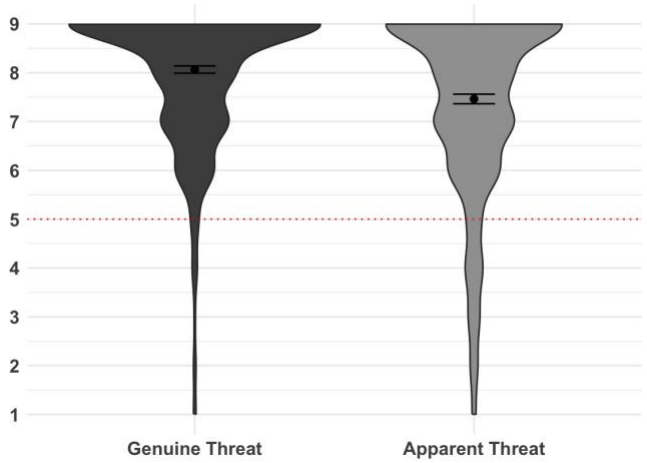

Figure 3. Violin plot for permissibility ratings in **Study 2**. Error bars show the standard error of the mean.

As with our previous studies, these results provide support for Subjectivism and ICV, but challenge Objectivism.

General Discussion

This paper explored a specific question in normative ethics, namely, under what conditions killing in self-defense is morally permissible. We distinguished two main positions: Objectivism and Subjectivism. Our results demonstrate that laypeople do not think that whether a threat is genuine or only apparent is as decisive as many Objectivists have argued. In contrast, the empirical evidence clearly aligns with Subjectivism and ICV, two positions that consider killings of apparent threats to sometimes be morally permissible.

However, we are hesitant to draw any conclusion as to whether Subjectivism or ICV has the upper hand. The studies revealed mixed evidence. In contrast to what ICV holds, we do not find a significant effect of Threat on permissibility ratings in concrete scenarios when using a between-subjects design. We do, however, find a significant effect of Threat on permissibility in concrete scenarios when using a within-subjects design, as well as when presenting participants with an abstract scenario in a between-subjects design.

That being said, while we did use three different stories and three different experimental designs (abstract scenarios tested between subjects, concrete scenarios tested between subjects, and concrete scenarios tested within subjects), we only manipulated whether the threat was genuine or apparent. There are more variations of self-defense cases relevant to the philosophical discussion at hand. We would like to discuss some of these variations, address a possible limitation of our current experimental and suggest empirical ways ahead.

Reasonability, Urgency, and Distress

In all our apparent threat scenarios, the defender's belief that he faces a threat is, while false, still reasonable. This resembles the apparent threat cases commonly used in the philosophical literature, as most Subjectivists claim that it is only permissible to kill apparent threats when it is *reasonable* for the defender, given his evidence, to believe they pose a threat to him.[6]

However, someone could object that most real-life cases of self-defense are epistemically much muddier; the defender is not merely plagued by uncertainty but also distress, urgency, and panic. He might not have the time or composure to deliberate on his evidence, and consequently, his belief that he is in imminent danger and has to defend himself might be unreasonable, even if the belief is psychologically inevitable given the distressing situation. One might argue that cases like these represent the typical real-life self-defense case more accurately than the typical philosophical thought experiments, which generally leave out any remarks about the defender being psychologically uncomposed or panicky.

These remarks raise a more general question for the empirical research conducted here as well as for non-experimental philosophy of self-defense: How does the reasonability of the defender's beliefs affect the permissibility of his killing in self-defense against merely apparent threats?

Most subjectivists—and all objectivists—will claim that it is impermissible to kill someone posing an apparent threat based on an unreasonable belief that he or she poses a genuine threat. Put differently, if a defender enters fight-or-flight mode and harms the apparent aggressor based on a belief that is not backed by his evidence, then killing in self-defense would not be permissible.

However, to date, there is no empirical evidence showing the impact of the reasonability of the agent's beliefs on laypeople's judgments of permissibility for killing in self-defense against apparent threats. A quick look at our Studies 1a and 1b reveals moderate to strong correlations between permissibility and reasonability judgments (all ps $< .001$; all rs $> .400$). In other words, the more reasonable laypeople thought it was for Defender to believe that his life was in danger, the more likely they were to judge the killing in self-defense as permissible. However, to get robust evidence on the relationship between the permissibility of killing apparent threats in self-defense and reasonability, more research is needed.

Philosophical Concepts and Folk-Conflations

The philosophical literature on self-defense utilizes a rich set of philosophical concepts with subtle distinctions. Specifically, distinguishing 'permission' from 'excuse' is vital for Objectivists, as this enables them to accommodate the intuition that, even though killing apparent threats is impermissible, Defender would not merit blame for doing so when his evidence suggests he is really in danger. However, one could question whether laypeople have this distinction. Can our participants simultaneously hold that an action is impermissible but that the agent is blameless? If not, perhaps

[6] See, e.g., Frowe (2010). One rare exception to this is Ferzan (2005).

the Objectivist could argue our current experimental setup 'rigs the game' in favor of Subjectivism.[7]

However, there is something to be said to lessen this worry. First, we specifically ask participants to evaluate both action-permissibility and agent-blameworthiness in order to avoid responses conflating these judgments. Second, in a pilot study, we asked participants whether they thought the permissibility- and the blame-question were distinct or roughly equivalent, and the (qualitative) responses showed no sign of conflation.

Of course, while this suggests that laypeople can acknowledge the distinction, it is still possible that they do not draw it in exactly the way Objectivists do. This itself need not be objectionable, though, as the disagreement between Objectivism and Subjectivism is a disagreement about how to draw the distinction, in terms of whether false beliefs can serve to *permit* actions or merely serve to *excuse* agents. If laypeople draw the distinction closer to how Subjectivism draws it, this would merely be another indicator that common opinion supports Subjectivism over Objectivism.

However, it would be fruitful for future research to look into how folk and philosophical concepts of 'permission' and 'excuse' compare. This is, however, out of scope for this paper.[8]

Exploring the Limits of Subjectivism

We believe our results tie into a larger debate about how morality relates to facts and appearances. While our cases manipulate whether the aggressor's gun is loaded or unloaded, the aggressor is, in all scenarios, pointing it at the defender and declaring his intentions to kill him. Regardless of whether or not the gun is loaded, the aggressor is doing something morally wrong here, and it is his own wrong behavior which makes the defender believe that his life is threatened and makes him shoot the aggressor. Thus, there is a clear sense in which it is the aggressor's 'own fault' if the defender kills him even if the gun is unloaded.

One might speculate whether lay judgments would change if this were not the case. That is, assume our cases involved a defender shooting an apparent threat because he reasonably believes they pose a threat to him, but where the apparent aggressor is *neither* responsible nor at fault for giving the defender this belief.[9] E.g., assume Defender has a violent stalker and that Defender, upon seeing someone running towards her in the forest at night, shoots the person, believing it to be the stalker.[10] In this case, Defender's belief that she has to kill the person might (depending on how we flesh out the details) be reasonable, but the person being harmed is not responsible for giving Defender this belief. If laypeople judge that killing the person is permissible, almost all ethical theories of self-defense would disagree with common opinion.[11] Thus, our research is part of a larger debate about whether morality in general requires that we do what is morally permissible according to our evidence or according to the facts.[12]

Acknowledgments

The research of Jan Garcia Olier and Pascale Willemsen were funded by the Swiss National Science Foundation (SNSF), grant number PZ00P1_201737.

References

Bolinger, R. J. (2021). The Moral Grounds of Reasonably Mistaken Self-defense. *Philosophy and Phenomenological Research,* (103), 140-156.

Coons, C., & Weber, M. (2016). *The Ethics of Self-Defense.* New York: Oxford University Press.

Doggett, T. (2011). Recent Work on the Ethics of Self-Defense. *Philosophy Compass*, 6/4, 220-233.

Ferzan, K. (2005). Justifying self-defense. *Law and Philosophy, 24*(6), 711–749.

Ferzan, K. (2012). Culpable aggression: The basis for moral liability to defensive killing. *Ohio State Journal of Criminal Law, 9*, 669–697.

Frowe, H. (2010). A practical account of self-defense. *Law and Philosophy, 29*(3), 245–272.

Frowe, H. (2014). *Defensive killing*. Oxford University Press.

Frowe, H., & Parry, J. (2022). Self-defense. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2022 Edition)*. Retrieved from https://plato.stanford.edu/archives/sum2022/entries/self-defense/.

McMahan, J. (2005). Self-defense and culpability. *Law and Philosophy, 24*(6), 751–774.

McMahan, J. (2011). Who is Morally Liable to be Killed in War? *Analysis*, 71, 544-559.

Nichols, S., & Knobe, J.. (2007). Moral Responsibility and determinism: the Cognitive Science of Folk Intuitions. *Nous*, 41(4), 663-685.

Otsuka, M. (1994). Killing the innocent in self-defense. *Philosophy & Public Affairs, 23*(1), 74–94.

Parfit, D. (2011). *On what matters (Vol. 1)*. Oxford University Press.

Quong, J. (2020). *The morality of defensive force*. Oxford University Press.

Thomson, J. J. (1991). Self-defense. *Philosophy & Public Affairs*, 20(4), 283-310.

[7] We are grateful to an anonymous referee for raising this point.

[8] However, as philosophers also disagree about how to draw the distinction, this raises the question of which philosophical conceptions such a comparison should focus on. Clearly, if it merely focuses on those used by Objectivists, it runs the risk of rigging the game in favor of them.

[9] This would be a case where the Responsibility Condition (mentioned in note 2) would not be met.

[10] Or, less realistically, assume Defender shoots the stalker's identical twin, believing the innocent twin to be the homicidal stalker.

[11] While Frowe (2010) deems Defender's killing permissible self-defense, almost all other theories deem it blameless wrongdoing (see, e.g., McMahan (2005), Otsuka (1994), Ferzan (2012), Frowe (2014), Quong (2020)).

[12] See, e.g., Parfit 2011: 150-151.