Penultimate draft. Please refer to the published version in The Journal of Political Philosophy for citation purposes.

Reasonable Mistakes and Regulative Norms: Racial Bias in Defensive Harm

Renee Jorgensen Bolinger*
Philosophy, University of Southern California

Edmund Joubran was working as clerk in a corner store when two men entered, one pointed a gun at him, and demanded cash. Believing they might kill him, Joubran drew and fired his own 0.38-caliber pistol, killing one of the men. Police later discovered that the robber's weapon was just a realistic toy gun.[1] Though he was mistaken about whether defense was necessary, it seems that Joubran did exactly what rational, well-intentioned agents in situations epistemically like his should be expected and permitted to do: protect their lives from an immediate, unjust threat. Contrast Joubran with another mistaken defender. Anthony Simon had a bad relationship with his neighbor Steffen Wong, and one evening the two got into a verbal argument. Simon believed that Wong was a martial artist and a lethal threat even empty-handed. When Wong appeared angry, Simon feared he was about to attack and defensively shot Wong. In fact Wong posed no threat, and did not even know any martial arts—Simon only believed that he did because he was Asian.[2] Intuitively, we should want legal principles for self-defense that count Joubran as having done nothing wrong, but hold Simon responsible for his unreasonable error. Unlike the robber, Wong can legitimately complain that his rights against harm have been violated.

We thus have some reason to adopt a *regulative norm* for self-defense that allows some reasonable mistakes, but recognizes that some victims of honest mistakes—especially when bias-driven—have legitimate cause for complaint. The function of this type of norm is to distinguish the conditions under which we will hold defenders to be innocent of any wrongdoing from those in which we hold them responsible for assault or manslaughter. Determining that an agent acted wrongly does not yet imply anything about whether she should be blamed or punished; the distinction serves only to demarcate cases where the victim of defensive harm has no complaint from those in which they have been wronged by the defender.[3] The norm must strike a fair balance between defenders' security, on the one hand, and

*My thanks to Jonathan Quong for his comments on several drafts of this paper. Thanks also to Rima Basu, Robin Jeshion, Richard John, Suzy Killmister, Mark Schroeder, Olúfémi O. Táíwò and audience members at the UConn Political Violence workshop in 2015 for fruitful discussion of many of the ideas presented here. I am indebted as well to two anonymous referees at this journal, whose constructive comments greatly improved the paper.

[1] Ridley (2012).

[2] This case is *State v. Simon* 646 P.2d 1119 (Kan. 1982). The jury was instructed that "A person is justified in the use of force to defend himself against an aggressor's imminent use of unlawful force to the extent it appears reasonable to him under the circumstances then existing", and acquitted Simon on the grounds that, given his beliefs, it was reasonable to defend himself. The Kansas Supreme Court later overturned this ruling, holding that the jury had been badly instructed and that the standard for reasonable belief requires both belief and the existence of facts that would persuade a reasonable man.

[3] There are many potential excusing factors that are not intuitively justifying; among them are coercion, extreme stress or duress, and non-culpable ignorance. An agent who is subject to one of these conditions may be fully excused (and thus not an appropriate target for punishment) but nevertheless have acted wrongly, violating the other party's rights against harm.

other agents' legitimate claim to live without fear of suffering mistaken defensive harm, on the other. While the socially adopted regulative norms need not perfectly track first-order moral determinations of when defensive harm is morally permitted, it is important that the norm be a *just* way of adjudicating the potentially conflicting rights claims of individuals in the society.[4]

The project of this paper is to trace how the demands of justice constrain the choice of regulative norm in a society with widespread group-directed bias.[5] My discussion will focus on the conditions in the United States, though the moral conclusions should generalize to societies with similar background conditions. I'll argue in Section II that we have reason to prefer a regulative norm that counts a mistake as reasonable when it was highly likely *on the evidence* that defense was proportionate and necessary to avert a threat. However, adopting an evidentialist norm under non-ideal conditions is treacherous business, and it is difficult to ensure that the fairness of the norm is not compromised by widespread bias. In Sections III-V, I'll briefly survey empirical data suggesting that the type and extent of bias prevalent in the US renders a straightforward evidentialist norm unjust, and since the legal practice in the US relies on such a norm, we must explore avenues for reform. Preferably this will take the form of adopting a modified evidential norm (I explore some options in Section VI), but if this proves impossible then we may have to accept a strict regulative norm instead, as the sole just alternative.

I The criteria for a good regulative norm

We generally take it that agents have rights against suffering defensive harm unless they have done something to make themselves peculiarly vulnerable to it. This reflects a commitment to the formal constraint that any permissible norm must

1. Afford each agent fair opportunity to avoid being subject to defensive harm.

If a norm fails this formal constraint it is internally unjust: by treating some as having less stringent rights against harm than others, it fails to respect the moral equality of the members of the society. Whether a norm satisfies (1) is a narrow question about the formulation, and can be evaluated without reference to the broader context in which the norm is adopted.[6]

As with other contexts in which fair opportunity is important, *merely* formal fairness does not ensure that agents enjoy substantive fair opportunities. Norms that would be just assuming full compliance, unbiased application, etc., may be disastrously unjust once we lift those assumptions. Given that the project is to identify a just norm for the regulation of defensive harm in the *actual* conditions, we should

[4] Here's one way fair regulatory norms can diverge from the first-order moral facts: suppose that culpably posing a merely apparent threat is not sufficient for an agent to forfeit their rights against harm. Then the robber retains his rights, and Joubran violates (or at least infringes) these rights by engaging in mistaken defense. However, if the robber has lost standing to enforce his right (either through counter-defense or by insisting on compensation), then it is not unfair for the regulatory norm to permit Joubran to impose defensive harm without facing legal consequences.

[5] Cognitive bias that disproportionately represents members of a given ethnic or racial group as threatening is the sort of group-directed bias that will likely compromise the justice in application of an evidential norm.

[6] This constraint is structurally parallel to the Rawlsian requirement that positions be open to talents, which requires only that the statement of the norms not directly mention groups, or use names or rigid designators to pick out some individuals for differential treatment. Also like the Rawlsian requirement, it must be coupled with constraints ensuring substantive fair opportunity before we can be confident in the substantive justice of norms satisfying the constraints. (See Rawls (1999), pp. 51-63.)

supplement (1) with substantive constraints that will jointly rule out norms that go badly awry as implemented under non-ideal conditions:

2. Enable agents to predict what behaviors will expose them to defensive harm.
3. Be applied equitably.

(2) ensures that there is adequate space for agents to exercise agency in determining whether to act in ways that will make them subject to defensive harm. If S can predict that ϕing will expose her to defensive harm, then she can intentionally avoid ϕing in order to avoid such exposure. Constraint (3) requires that the fair opportunities afforded by any norm that satisfies the structural constraint be preserved in implementation.

II Candidate Norms

A norm's claim to fairness depends in part on the closeness of the match between its verdicts and the prescriptions of a moral theory for the domain. Without attempting here to settle which is the best theory of the moral permissibility of self-defense, we can get a rough sense of some of the prospects for a norm by noting which theor(ies) it must invoke, and what idealizing assumptions it must make.

A Strict Norms

A strict norm counts defensive force justified only if it is in fact proportionate and necessary to avert an unjust threat to the defender (or fact-relative morally justified)[7], and is perhaps most obviously even-handed in form. It satisfies the formal constraint easily: each agent (at least in principle) has the opportunity to avoid willingly posing an unjust threat to others' rights. It is also usually easy to anticipate and avoid behaviors that would pose such threats, satisfying (2). If the norm is applied at least somewhat conscientiously, it will also satisfy (3): the strict norm is harsh to mistaken defenders, but it is uniformly harsh. If the threat is merely apparent, then no matter what her evidence was, the mistaken defender is has acted impermissibly (though perhaps excusably). Strict norms are thus defensible—their verdicts correspond closely to those of plausible moral theories, and they satisfy our three constraints on substantively just norms. But they aren't the *preferred* choice, since they lack the flexibility to allow that reasonable mistakes are at least sometimes permissible.

B Individualized Evidentialist Norms

A natural way for a norm to make space for reasonable mistakes is to hold that defenders act permissibly when they *believed* that defensive harm was fact-relative justified. For this to be plausible we must idealize a bit. The idea is not that permissibility rests simply on what beliefs or evidence a particular agent *happens* to have. Instead, it depends "on what it is reasonable for the agent to believe in the situation, what it is reasonable for the agent to do to check those beliefs, and whether the agent has

[7] For an incomplete but representative sample of fact-relative views of the moral permissibility of defensive harm: on Thomson (1991)'s account, one incurs liability to defensive harm by being causally responsible in some way for an unjust threat to another's rights. McMahan (2009) and Otsuka (1994) defend a narrower variant of roughly the same structure, on which S is liable when she is morally responsible for an unjust threat facing the defensive agent. These views can (and many do) hold that a reasonably mistaken defender acts in an evidence-relative permissible way; that is, her action would be morally permissible *if the facts were as her evidence suggests*. But on these accounts it is fact-relative permissibility that determines whether the act was wrong; evidence-relative permissibility serves only to determine whether an agent is culpable or blameworthy for her mistakes.

done those things."[8] The evidentialist norm can be developed either as an *individualized* standard, or as a shared *public* one; I'll discuss the individualized version first.

Individualized evidential norms count an agent justified in imposing defensive harm just if, conditional on all the agent's (relevantly idealized) beliefs,[9] the probability that defense is necessary exceeds the relevant likelihood threshold.[10] The idealization precludes any beliefs or evidence that an agent has only because she is a bad epistemic agent (beliefs resulting from bias, motivated reasoning or culpably bad evidence-gathering) from grounding *reasonable* belief in the need for defense. So the norm satisfies constraint (1): all well-intentioned agents can in principle avoid acting in a way that makes it reasonable for another individual to believe they are an aggressor.

However, there are two problems with an individualized standard. First, because it ties reasonableness to the *individual's* set of evidence, the norm struggles to meet condition (2): it does not put agents in a position to predict and avoid risk-incurring behaviors. Without knowledge of the defender's other beliefs and experiences, it is impossible to anticipate all of the behaviors that will make it likely *given her other evidence* that you are an aggressor. The idealizations do little to help: they only disqualify the agent's defective beliefs; they do not grant you access to her total (remaining) evidence. Certainly you can avoid behaviors (like aiming a weapon at someone's head) that are sufficient to ground such a belief independently of other evidence. But you remain unable to know whether, e.g., wearing a Lakers jersey and approaching from the north at 6 p.m. makes it likely *on the other agent's evidence* that you are an aggressor. It thus leaves you little space to intentionally avoid behaving in a way that makes it reasonable to think you are an aggressor, because it does not help predict which behaviors need avoiding.

The second problem is that if adopted against a background of prevalent group-directed bias, there is real risk that *as implemented* the norm will violate the substantive even application condition (3). To preserve fair opportunity, we must hold agents like Simon—whose acts were reasonable *only on their non-idealized evidence*—responsible for their errors. To do that, we need to be able to recognize when an agent's beliefs fall short of the relevant idealizing evidential standard. We can't expect defenders to successfully self-regulate: while *consciously prejudiced* defenders might be aware that their beliefs aren't reasonable, merely biased defenders typically aren't.[11] Agents are more attentive to threat-indicating

[8] Scanlon (2008), p. 52.

[9] Strictly speaking, we actually conditionalize on the agent's full epistemic partition: a statespace that assigns probabilities to every proposition, reflecting the agent's credence in that proposition being true. Only the propositions in which the agent has high or resilient credences are properly called 'beliefs'. However, for the purposes of this paper, I will gloss over all of these subtleties and simply use 'beliefs' in a loose, inclusive way.

[10] In the literature on the moral permissibility of self-defense, several authors have offered evidential accounts that would vindicate a norm of this sort. Frowe (2010) argues that a defender is morally justified in imposing defensive harm if it was reasonable for her to believe that doing so was necessary and proportionate to avert an unjust threat. Uniacke (1994) develops a similar account of subjective permissibility, but preserves an actual-threat aversion account of objective permissibility. Lazar (2014) defends a similar view in decision theoretic terms: defenders are subjectively permitted to impose defensive harm when doing so is the option that maximizes expected value, on the evidence an epistemically virtuous agent would have and probabilities she would assign in the defender's circumstances. Ferzan (2012)'s proposal that defenders are justified if (i) they believe that defensive force is necessary, and (ii) the apparent aggressor is culpably responsible for this belief can also be glossed as an evidential view, though the second condition is objective rather than epistemic, rendering the view a hybrid.

[11] I'll use 'conscious prejudice' to refer to consciously held or explicit acceptance of attitudes or beliefs denigrating the target group. This is contrasted with (implicit) 'bias', which refers to patterns of thought, attention, and conceptual associations that operate sub-consciously and influence agential behavior in inegalitarian ways, sometimes in direct conflict with the agent's consciously endorsed egalitarian views.

evidence about persons who match their heuristic for threateningness. If the heuristic encodes a group-directed bias, her disproportionate attention will translate into disproportionate evidence that a member of that group is an aggressor. Unless the agent is aware of her bias, this belief will seem to her to be as evidentially well-supported and respectable as they come.

If the bias is especially widespread, we may not correct defenders' errors at the judicial level, either. When asked to make judgments about whether what an agent believed was reasonable, given the circumstances, judges and juries must consult their own sense of the situation. They may do their best to decide cases fairly, but good intentions are not sufficient to idealize away the effects of bias: jurors are members of the broader population, and are as susceptible to adopting faulty heuristics as anyone else. If the jurors are subject to the same subconscious group-directed bias as the defender, they are likely to misjudge the evidence in the same way that the defender did. So if the mistaken defender exhibited biases foreign to the jury, she is likely to be held to account for them, but if jurors rely on the same error-generating heuristics, they will find the agent to have believed and acted reasonably.

An individualized evidential norm is therefore doubly problematic: it doesn't reliably put agents in a position to anticipate what behaviors will expose them to defensive harm, and if there is widespread group-directed bias then unless the norm is carefully constrained, it will fail the even application condition.

C Public Evidentialist Norms

Rather than exonerating agents based on the totality of the individual's idealized evidence, *public evidential norms* privilege a set of publically available evidence and hold that agents act permissibly when, on that privileged evidence, it is sufficiently likely that defensive harm is justified. I develop a particular account of this type in detail elsewhere,[12] but here will simply sketch the general outline as an alternative to the individualized accounts. There are two reasons why this is a valuable exercise: first to illustrate the breadth of available evidentialist norms, and second because it makes some options for reform easier to state. The individualized model tempts us to think of reform as a project of rooting out internalized, individual bias; a project not immediately amenable to social solutions.[13] Because a public evidence approach emphasizes the role of society in determining what evidence is privileged, it helpfully focuses attention on how social decisions can help ameliorate the effects of bias.

Under a public evidential norm, a society can treat some behaviors as marked 'signals' of aggression, licensing agents to assume (absent countervailing evidence) that the performer is an aggressor.[14] If the

[12] See Bolinger (ms).

[13] Emphasis on individual bias can lead us to think that the problem is located and must be addressed at a radically personal level, or else that it can only be addressed by sweeping reforms of the kind proposed by Anderson (2010), requiring thorough-going racial integration in neighborhoods, workplaces, etc. Mid-level reforms to immediately address the most severe consequences of bias are apt to get lost in the shuffle when the problem is approached from this perspective, but are easier to see on public evidentialist accounts.

[14] The notion of signals that I rely on here roughly follows Skyrms (2010) in treating them as properties (which may range from phenotype expression to the intentional performance of a specific behavior) that become associated with contents by the combined effects of reinforcing effective signals and forgetting unsuccessful ones. When a behavior b strongly correlates with some state of affairs X, and reacting to b by ϕing yields a rewarding outcome for the receiver, the receiver becomes more likely to take b as a sign that X, to react to b by ϕing in the future, and to tell other members of the community to do so as well. One can unintentionally send a signal by doing something as simple as *instantiating* a conventionally significant property. For example,

signals are public knowledge and it is reasonably easy for members of the community to avoid the signaling behaviors, then this has the effect of establishing a set behaviors for which the apparent aggressor can be held responsible. Like the robber in the opening anecdote, agents who perform these behaviors have no complaint against mistaken defenders, since they induced the error (whether intentionally or not) through their own easily avoidable behavior.

As long as the signals are in-principle avoidable, this norm formally affords all agents fair opportunity to avoid incurring vulnerabilities, so it satisfies constraint (1). It also co-ordinates what will be used as heuristics for aggression, allowing agents to avoid the marked behaviors, thus satisfying the second constraint. However, like the individualized standard, it is still susceptible to distortion by widespread bias. A signal can be established just by sufficiently widespread use, and it is well within the power of a biased community to use markers like race as signals of threateningness. Such signals are *perverse*: they fail to afford some subgroup (members of a particular race) fair opportunity to avoid appearing to be aggressors. So, like the individualized epistemic norm, the public evidential norm can be rendered unjust in conditions of widespread bias.

D Actual Practices

The fact that the fairness of straightforward evidential regulative norms is threatened by prevalent group-directed bias has more than merely theoretical significance. We do not simply have an abstract preference for regulative norms that exonerate agents for reasonable mistakes; they are also embedded in English and American accountability practices regarding defensive harm. We systematically exonerate agents who impose defensive harm in situations that we think a reasonable person would find threatening, and display little sympathy for victims who, intentionally or not, gave mistaken defenders good reason to act.

Many states incorporate a 'reasonable belief' standard into the conditions for self-defense, rather than requiring that the defendant have actually faced a genuine threat. For a representative sample: Washington uses a subjective standard, "requir[ing] the court and the jury [to] evaluate the reasonableness of the defendant's *perception* of the imminence of danger in light of all the facts and circumstances known to the defendant." In New Jersey, "The use of force against another in self-defense is justifiable when the actor *reasonably believes* that such force is immediately necessary for the purpose of protecting himself against the use of unlawful force." Arizona statute provides that "a person is justified in using physical force against another when and to the extent that *a reasonable person would believe* that physical force is immediately necessary to protect oneself", and Colorado allows that a person "may use a degree of force which he reasonably believes to be necessary" in self-defense.[15]

The proper interpretation of the reasonableness standard for self-defense in law is a matter of lively debate.[16] The epistemic reading prominent in English caselaw suggests an individualized evidential

if you wear a red polo shirt and khakis to Target, you signal that you are a Target staff member, even if you had no intention of doing so.

[15] *State v. Janes*, 822 P.2d 1238, 1242 (Wash. Ct. App. 1992), citing State v. *Wanrow*, 559 P.2d 548, 556 (Wash. 1977)); *State v. Kelly*, 478 A.2d 364, 373 (N.J. 1984); *State v. Romley*, 836 P.2d 445, 451 (Ariz. Ct. App. 1992); *People v. Miller*, 981 P.2d 654, 658 (Colo. Ct. App. 1998). Emphasis added.

[16] There are roughly three positions in debate over the proper interpretation of the 'reasonable belief' standard. Some advocate

construal, while the more normatively laden interpretation common in many U.S. courts suggests a more public evidential standard. Either way, in counting defenders to have acted permissibly so long as their mistakes were "reasonable," our practices must rely on some type of evidential model of justified self-defense. Since unconstrained evidential norms like these are likely to be unjust if actual conditions in the society exhibit high levels of group-directed bias, the immediate next question must be "is our society sufficiently free of these compromising influences?"

III Present conditions: racial bias in defensive decisions

Our explicitly endorsed practices are not obviously impermissible: justifications that make direct reference to racial, ethnic or classist group sortals are rarely offered, and accepted even less frequently.[17] However, the absence of obvious injustice does not mean that our practices satisfy the requisite moral standards. Particularly in a society that at least nominally rejects racist or classist discrimination, agents are motivated to appeal to such markers only obliquely; instead, they are likely to report simply that "he was threatening" or "I feared for my life". Still, it is possible to ferret out specific heuristic practices by tracking agent behavior.

A Threat perception in experimental conditions

The *weapons bias test* is designed to test for differential response times and patterns to individuals of different races. Participants must identify a pictured object by pushing one of two buttons, 'shoot' for weapon and 'don't shoot' for a non-weapon. In one variant, objects are shown immediately after a picture of a face; in the more common variant, the objects are held by pictured individuals.[18] All pictured individuals are adult males, roughly half are white and half Black. Participants are instructed to ignore the people and just focus on weapon identification. When response windows were generous (860 ms), most participants' final determinations were accurate, but under more restrictive time limits (630 ms) they made patterned errors:

> The results of some 20 studies with this task, with a variety of parameters and manipulations, consistently show racial bias in both the speed and accuracy with which such decisions can be made. Participants are faster and more accurate when shooting an armed Black man rather than an armed White man, and faster and more accurate when responding "don't shoot" to an unarmed White man rather than an unarmed Black man.[19]

reading it as a *subjective* requirement that the belief be sincere, others as invoking a *typical* person, and still others (including Lee, 2003; Armour, 1994) as a normative standard for what the agent *should* believe. This particular debate is orthogonal to our concerns at the moment, as the only standard of interest to our discussion is a normative one. The subjective and typicality interpretations are transparently subject to distortion by bias, while the normative standard may be more resilient. For further discussion on the reasonable person standard, see Zipursky (2015).

[17] Of course, the fact that a consideration is not explicitly raised in a court room is not evidence that it plays no role. The trial of *People v. Goetz* is infamous for the implicit use of race as a justification for 'reasonable' self-defense. Stephen Goetz (a white adult male) was approached on a New York subway by four Black youths, who asked him for five dollars. Goetz shot the youths, subsequently claiming self-defense. The issue of race was not explicitly broached at the trial, but the Defense had young black men 're-enact' the scene for the jury, which many scholars have condemned as implicitly feeding on jurors' racial anxieties to secure sympathy for Goetz. (For more thorough discussion see Fletcher (1988, pp. 206-8)).

[18] Studies following this design were run by PaynePayne (2001); Correll et al. (2002); Greenwald et al. (2003); Plant et al. (2005); Correll et al. (2007); Kahn and Davies (2010); Ma and Correll (2011); Correll et al. (2011) and Sim et al. (2013), among others. The bulk of studies on this effect have focused on white/Black bias in the United States, so the discussion to follow will also focus narrowly on this subtype of group-directed bias.

[19] Correll et al. (2007, p.1007)

A follow-up study sorted targets into three categories: highly phenotype-stereotypical Black (HS), low phenotype-stereotypical Black (LS), and white. Error rates remained pronounced for HS targets, but were not significant for either the white or LS targets.[20]

There are two plausible explanations: either participants had greater difficulty distinguishing weapons from non-weapons when the target was black (discrimination error), or they required less evidence to feel sure enough to 'shoot' under that condition (lower criterion). Based on the patterns of error and delay times, theorists concluded that "participants set a lower decision criterion to shoot for African American targets than for Whites. That is, if a target was African American, participants generally required less certainty that he was, in fact, holding a gun before they decided to shoot him."[21] Further studies found that so long as the task is simple, the details of the errors match what we should expect from a low criterion. However, when the task is more complex, subjects also display difficulty distinguishing threatening from non-threatening objects held by Black targets, making discrimination errors that compound the effect of the lower criterion.[22]

Importantly the asymmetry in criterion isn't driven by conscious prejudice, and displayed bias was as strong among minority and Black participants as among white participants. Subjects' responses to target blackness was parallel to their response to contextual cues of threateningness, and the asymmetry between their willingness to shoot Black vs white targets disappeared when some other contextual cue indicated that the white targets were dangerous.[23]

A heuristic may also function indirectly as an aid in interpreting ambiguous information. For instance, if Polly notices a person is walking behind her at night, she may rely on a signal to determine whether what she sees is dangerous (someone stalking her) or not (merely coincidentally going the same way). We are rarely consciously aware of the heuristics we use to resolve ambiguities in this way, and when a heuristic is firmly established we resolve ambiguity so quickly that we no longer perceive it as having been ambiguous. If the signal cuts along group lines, the very same behavior will be interpreted as threatening when performed by a group member, but not when performed by others. Racial cues appear to perform this function as well. Study participants perceived the same behavior as threatening when performed by a Black agent (classifying it as a "violent shove"), but non-threatening when performed by a white agent (classifying it as a "light push"). Both white and African-American participants exhibited the same perception bias.[24]

The tendency to resolve ambiguous information in favor of threateningness when the agent is black has worrisome implications for a reasonable belief standard of self-defense. All else equal, defenders are more likely to believe that they face a threat if the agent is black. Explaining their findings, Correll et al. (2002) suggest that "a participant who catches a glimpse of some elongated shape in the target's hand

[20] Kahn and Davies (2010). Targets in the HS group were selected for looking 'stereotypically Black': having dark skin, broad nose, and full lips. Kahn and Davies' results are corroborated by Ma and Correll (2011).

[21] Correll et al. (2002, p. 1325.) This is consistent with the findings of Correll et al. (2007), Kahn and Davies (2010), Correll et al. (2011) and Sim et al. (2013).

[22] Greenwald et al. (2003); Klauer and Voss (2008).

[23] Correll et al. (2011, p. 187).

[24] Duncan (1976); Sagar and Schofield (1980).

may draw on stereotypic associations, interpreting the shape as a gun if the target is African American but as a cellphone if he is White." In fact, there is some evidence that the apparent race of the target has a bi-directional relation to subjects' ability to perceive crime-related objects. A series of studies by Jennifer Eberhardt found that"Black faces triggered a form of racialized seeing that facilitated the processing of crime-relevant objects, regardless of individual differences in racial attitudes," and "White faces inhibited the detection of crime-relevant objects."[25]

B Perverse Signals and Citizen Self-Defense

Attentional bias of this kind operates independently of conscious prejudice, and would be expected to have the dual effect of both making group members more likely to suffer mistaken defense, and making it more likely that their assailant will be exonerated. It is near-impossible to test this expectation against meaningful distributional data from archival records of actual cases. When self-defense is considered justified, cases often don't go to trial and so are only reflected in sparse datasets like the Federal Bureau of Investigations' Supplementary Homicide Report (aggregating data for homicides from 2005 to 2010). In analyzing this dataset, John Roman found that

> . . . controlling for all other case attributes, the odds a white-on-black homicide is found justified is 281 percent greater than the odds a white-on-white homicide is found justified. By contrast, black-on-white homicide has barely half the odds of being ruled justifiable relative to white-on-white homicides. Statistically, black-on-black homicides have the same odds of being ruled justifiable as white-on-white homicides.[26]

These disparities were noticeably greater in states with Stand Your Ground laws. While striking, this data is not especially robust: while the dataset includes information about whether law-enforcement considered each homicide 'justified', it does not include relevant additional information about the circumstances of each homicide, so Roman was unable to control for factors like location (i.e. street or residence) that may legitimately affect whether a homicide is considered justified.

Simulated jury proceedings can be much more carefully controlled, and the data from a series of experiments by Sommers and Ellsworth suggest that group-directed bias *does* affect jurors' opinions of whether a defensive shooting was justified. They provided participants with a written summary of a domestic assault case, giving half the mock-jurors a summary describing the defendant as a Black male, and the other half a summary describing him as white. The jurors rated the prosecution's case against the Black defendant stronger than the case against the white, considered the arguments in his defense to be "significantly weaker than the White defendant's defense," and perceived the white defendant as "more honest and moral" and the Black defendant "more aggressive than the white defendant"—even though the cases and arguments were identical except for the race of the defendant.[27]

---

[25] Eberhardt et al. (2004, p. 877).

[26] Roman (2013, p. 9).

[27] Sommers and Ellsworth (2000, 2001). Interestingly, these discrepancies disappeared if race was made a focal point in the case by adding that the defendant had been the victim of racial slurs. Their findings are representative of the overall findings of a series of studies on the effects of bias in jury proceedings; for an excellent summary see Hunt (2015), esp. pp 276-277, 282.

C Perverse Signals and Police Self-Defense

Insofar as police officers are susceptible to biases present in the general population, we should expect that if race is a commonly used heuristic for threateningness, we will see evidence of that in the error rates of police use of defensive force. Again there are difficulties gathering adequately rich data, but available data matches these expectations. A database of police fatalities in the US in 2015 (managed by *the Guardian*) found that in the first five months, police were nearly twice as likely to make a false-positive error when the subject was black: "while 30.5% of black people killed were unarmed, 16.1% of white people killed had no weapon."[28] The discrepancies between these numbers dropped a little by the end of the year but remain alarmingly disproportionate; in the final count, 24.8% of black fatalities and 17.5% of white fatalities were unarmed.

Though the *Guardian* dataset includes only killings that occurred in 2015, it is consistent with the findings of Ross (2015)'s more robust analysis of police-involved shootings in the US between 2011 and 2014. Ross found that for individuals in a police encounter, "the median probability across counties of being {black, unarmed, and shot by police} is 3.49 times the probability of being {white, unarmed, and shot by police}."[29] Perhaps more telling, on average across all US counties, in an encounter with the police Black *unarmed* individuals are roughly as likely to be shot by the police as white *armed* individuals. Being armed made a significantly bigger difference for white individuals' risk of being shot by police than it did for black or hispanic individuals. Ross notes that "this pattern is consistent with police being more discerning of armed/unarmed status before shooting a white suspect than shooting a black or hispanic suspect." Importantly, these differences did *not* strongly correlate with differences in race-specific crime rates as measured by assault and weapons-related arrest rates.[30]

IV Are these heuristics really unfair?

All together, these observations strongly suggest that Black Americans are perceived as more threatening, all else equal, that this is widespread enough to distort jury judgments, and that it tends to increase their risk of suffering mistaken defensive harm. This is bad news for either type of evidential norm. Straightforward individualized norms are rendered unjust when bias is widespread enough to undermine even application of the norm, which the jury data suggest it is. The shared public norm is unjust

[28] Swaine et al. (2015), Swaine (2015). Individuals were counted as 'armed' if they held a knife, gun or gun replica, another implement likely to cause severe harm, or operated a vehicle as a weapon (by driving it at someone). They were considered unarmed if either they held no objects, or the object was unlikely to cause serious harm (such as a broom handle).

[29] Ross (2015, p. 6.)

[30] The fact pattern in high-profile errors suggests that police are more susceptible to discrimination errors—mistaking a wallet or other innocuous object for a deadly weapon—when facing a black suspect. Salient examples include the fatal shootings of Amadou Diallo, Anthony Dwain Lee, and Timothy Thomas. In 1999 Diallo was approached on his porch around 2am by four plainclothes NYPD officers, ordering him to show his hands and asking for ID. Diallo reached into his jacket and withdrew a wallet, which one of the officers mistook for a gun. The officers fired 42 shots, 19 of which hit and killed the unarmed Diallo. The officers were tried for second-degree murder, but were acquitted on a self-defense claim (?). In 2000 Anthony Dwain Lee was at an upscale Halloween party. LAPD officer Hopper responded to a noise complaint at the house, walked around the side and saw Lee through a window. Lee was in a costume including a replica pistol, in a room with two other costumed partyers. Hopper fired nine shots at Lee, four of which hit and killed Lee. The LAPD internal review board found that the shooting was "justified because the officer believed the gun was real and feared for his life." (Lait and Glover, 2001). In Cincinnati in 2001, Timothy Thomas had several outstanding warrants for non-moving traffic violations. Late one night, an officer approached him, and Thomas ran, ultimately turning down a dark alley. Another officer entered the alley and fired a single round, which struck Thomas in the chest, killing him. The officer later reported thinking that Thomas had a gun, but no weapon was found (Larson, 2004).

when the heuristic for threateningness is not adequately avoidable, and the data suggest that race is being used as such a heuristic. Before concluding that the operative regulatory norm is morally compromised, we should consider two objections to my claim that it is unjust to use racial heuristics for threateningness. I'll call them the *statistical* and *avoidance* arguments, respectively.

A The Statistical Argument and Default Signals

The statistical argument claims that it is not unjust to be more suspicious of a group, so long as the heightened suspicion is roughly proportionate to the demographic's higher probability of being involved in violent crime. Even setting aside concerns about the social causes and dubious probative value of such statistics, there are two reasons why this sort of 'proportionately disproportionate' suspicion is unjust. First, it is near-impossible in practice to keep such statistics from exerting an outsize influence on estimates of threateningness. People are quite bad at reasoning with probabilities, and tend to substitute resemblance to salient paradigm cases.[31] Additionally, insofar as racial disparities in crime-rates are driven by disparate poverty levels, allowing reliance on racial statistics encourages double-counting. Having already accounted for the higher probability of crime *given economic status*, race yields little to no additional data about an individual's likelihood of being an aggressor. The result in that agents systematically over-estimate how likely minorities are to be involved in crime.[32] So permitting reliance on statistical inference of this sort would violate the even application constraint (3) in practice.

Second, even assuming that agents could constrain their credences to perfectly match demographic statistics, such reasoning fails to respect the agency of members of marked minorites. Rather than being judged by their actions, they are instead assumed (in virtue of their group-membership) to be more threatening unless they actively *cancel* or undermine that assumption. This need not be *difficult* to be burdensome. In Claude Steele's *Whistling Vivaldi*, Brent Staples recounts his own cancelling behavior: while walking at night,

> I tried to be innocuous but didn't know how . . . I began to avoid people. I turned out of my way into side streets to spare them the sense that they were being stalked. . . Out of nervousness I began to whistle and discovered I was good at it. . . I whistled popular tunes from the Beatles and Vivaldi's *Four Seasons*. The tension drained from people's bodies when they heard me; a few even smiled as they passed me in the dark.[33]

An agent like Brent cannot think through a difficult logical proof on his way home, if he can't simultaneously whistle a complex classical tune and focus on derivations. He can't enjoy the silence. He can't listen to hip-hop through earbuds. He must also constantly monitor his behavior, lest an ordinarily inocuous activity thrust him past the threshold of 'sufficient threateningness.' In short, there are a wide range of opportunities open to other agents which he is forced to forgo or else risk suffering mistaken defensive harm, and the sole justification for this reduced space of alternatives is that he appears Black.[34]

---

[31] For more on this point, see the discussion in Richardson and Goff (2012).

[32] Armour (1994, p. 794) cites precisely these reasons in his argument against using racialized base-rate data in trials. He argues that jurors and defendants systematically over-estimate the likelihood that a Black agent is involved in criminal activity, so introduction of demographic statistics is likely to be "more prejudicial than probative."

[33] Brent Staples, as quoted in Steele (2011), p. 21.

[34] This sort of unjustified reduction of a space of alternatives for an agent, relative to other agents in the society, counts as a wrongful restriction of the agent's autonomy on both the picture developed by Raz (1986) and on the view defended by Steiner

In addition to these losses, marked minorities face increased risk of harm if they are unable to adequately cancel the signal. This plays out in police interactions, in jury verdicts, in the unwillingness of bystanders to offer assistance or extend the benefit of the doubt. As a case in point: in 2013, 24-year-old Jonathan Ferrell was in a terrible car accident in Charlotte, NC. After pulling himself from the wreckage, Ferrell knocked on the door of a nearby house, asking for help. Frightened, the resident called the police, who shot Ferrell 10 times on arrival, killing him.[35] Ferrell's fate illustrates a core problem with default signals for threateningness: even if *usually* easy to cancel, they expose the signaler to substantial risk when circumstances render them unable to effectively cancel the signal. The fact that their killer *might* later be held responsible for the mistake is cold comfort; the agent has a justice-based claim against being in put in a position where they appear threatening *by default*.

Thus the statistical argument fails: members of minority groups can legitimately claim that it is unjust to expose them by default to higher risks, grounded solely in frequency facts about a broad demographic group. So if an evidential norm exonerates mistaken agents who rely on statistics of this kind, it cannot provide minorities with adequate substantive opportunity to avoid harm. Such a norm therefore fails our constraints, and is unjust.

B The Avoidance Argument: Behavior-based Heuristics

The avoidance argument holds that while it would be unjust to rely on an unavoidable signal, the data just covered doesn't establish that *that's* what we're doing. In the experiments, only subjects who looked *stereotypically* Black suffered noticeably higher false-positive error rates. To avoid appearing threatening, perhaps all an agent needs to do is avoid dressing, speaking, or acting in stereotypical ways. Geraldo Rivera suggested something like this in the wake of Treyvon Martin's killing, advising Black parents to keep their kids from wearing hoodies.[36] This approach to avoidance is also known in some circles as 'Respectability Politics.'

I'll grant that, as Randall Kennedy argues, this can be an effective and rational policy for individuals to adopt, *given that* the broader culture exhibits the biases outlined earlier.[37] It is a further question whether the availability of such strategies suffices to render continued reliance on such heuristics permissible. To show that, we would need to demonstrate that members of the broader culture can reasonably demand of Black agents that they behave in the ways dictated by respectability politics or else accept heightened risks of suffering mistaken defensive harm. I contend that such a demand cannot be reasonably made. Though potentially low-effort, the avoidance behaviors required by respectability

(1994).

[35] Officer Kerrick was indicted and tried for killing Jonathan Ferrell, but the case resulted in a hung jury and the prosecutor's office declined a retrial (Katz, 2015). A similar case occurred a year later: after suffering injuries in a car accident, 19-year-old Renisha McBride went to a nearby house to ask for help. Instead of helping, the resident, 55-year-old Theodore Wafer, fatally shot McBride through a locked screendoor. Both shooters subsequently claimed self-defense, on the grounds that the victim had appeared agitated and threatening. Wafer was ultimately convicted of second-degree murder and manslaughter and sentenced to 17 years in prison, after a jury found his claim to have feared for his life unreasonable (Semuels, 2014).

[36] During the March 23 2012 show of *Fox & Friends*, Rivera said "I am urging the parents of black and Latino youngsters, particualrly, not to let their children go out wearing hoodies! I think the hoodie is as much responsible for Trayvon Martin's death as George Zimmerman was" (see http://www.politico.com/news/stories/0312/74392.html). For a good discussion of whether attire like hoodies or sagging pants are worth accommodating, see Jeffers (2013).

[37] Kennedy (2015).

politics involve giving up a particular cultural identity that is not in itself objectionable, which the agent is entitled to retain. The ability to select for oneself which goals to pursue and what to value is at the core of valuable exercise of agency; thus forcing an agent to give up something which she has decided to pursue counts heavily against adequate avoidability. While it is permissible for an agent to freely choose to sacrifice a project or cultural entitlement, it cannot be demanded of her. Costs of avoidance that either threaten a project of great importance to the agent unless she behaves in a certain way (effectively coercing her), or significantly reduce the range of options, undermine the agent's ability to autonomously shape her own life.

A formal opportunity to avoid appearing threatening is not adequate if we cannot reasonably demand the agent make the sacrifices it requires, so if an agent's only options for avoidance are all high-cost, then she has not been given *adequate* opportunity to avoid appearing threatening. So the Avoidance argument also fails: the heuristics in use are not adequately avoidable, so an evidential norm that exonerates agents for relying on them is unjust.

V So, What Does Justice Demand?

Since an evidential norm that is in practice shaped by racial heuristics is unjust, we have strong reason to reform. This may come at significant social cost. Reform requires that we refuse to exonerate agents who acted on perverse signals or biased heuristics. It is not enough to simply admonish agents to avoid acting on bias. Stable heuristics are points of cost/payoff equilibria: a perverse signal will persist as long as it is the most cognitively efficient and reliable means of distinguishing between cases where the defender will be socially permitted to act and those where she won't. The cost of false-negative errors is stably high: the attacker succeeds in inflicting grievous harm. The cost of false-positives depends on the strength of the victim's (or her representatives) position to demand compensation or punishment. Perverse signals persist precisely when the groups disadvantaged by the signal lack power to effect a corresponding shift in social practices of exoneration. Given that the costs of false-negatives will remain stable, we cannot expect agents to expend significant cognitive energy to find a new heuristic unless we change the costs of false-positives, effectively disrupting the signal equilibrium.

A The Moral Imperative of Reform

Even if the only way to excise the effects of racial bias were to adopt a strict regulative norm, the moral reasons are decisive. Exonerating agents for reasonable mistakes reduces overall odds that members of the society will suffer unjust harm, but the cost of securing this good is borne by the individuals who get reasonably mistaken for aggressors. If every member of the society had similar *ex-ante* chances of suffering the costs, this would just be a roulette problem, decidable by comparing the expected value of the benefit to the expected cost of the risk for each individual. But this is not the structure of the problem we face. The question we must answer is whether the benefits to one population, the A-group, are sufficient to justify the imposition of serious costs on a distinct population, the B-group. Perhaps it is possible in principle to justify policies of this structure, if there is enough of an asymmetry between the size of the populations, and the gains to A are great enough. But this sort of calculated sacrifice of the strong interests of a minority is the kind of a greater-good reasoning that most deontologists are concerned to avoid, at least unless there is such a notable asymmetry that it seems unconscionable to

preserve the rights of the few. It is doubtful that the case at hand comes anywhere close.

When the standards determining which mistakes are 'reasonable' are partially shaped by racial bias, use of a reasonable belief standard forces an already vulnerable minority to face increased risk of harm so that others can reduce their own risks. This is unjust in two distinguishable ways: (i) it imposes unfair risk of suffering serious harm on the minority, and (ii) it treats them as moral inferiors by refusing to acknowledge bias-driven harms to them as violations of their rights. This is morally unacceptable.

B The Preferred Option

The ideal solution is to reform the evidential standard for permissible mistakes: change the heuristics, but retain the general structure. Many scholars have offered a variety of suggestions for reform, ranging from race-flipping jury instructions to treating self-defense as a mitigating (excusing) rather than justifying circumstance in cases that weren't obviously reasonable and race was a factor.[38] These measures may be useful in recognizing wrongful errors at the judicial level, but full reform—reform that has a chance of changing the error rates—must also aim to undermine race-based public heuristics for threateningness.

There is some grounds for optimism here. The racialized signals aren't hardwired: participants rely on them only so long as they are a useful proxy for threateningness. In weapons bias tests, when the costs of false-positive errors are increased, or an alternative, more efficient proxy is introduced, subjects stop relying on the racialized signals. When the test environment randomizes the distribution of guns so that the race of the pictured individual does not correlate with the appropriateness of a "shoot" response, participants evidence a racial weapons bias only on the first round of testing. In subsequent rounds, they simply stop relying on race as a useful proxy. These lessons hold: when subjects are re-tested 24 hours later, they do not manifest a racial weapons bias.[39]

To replace racial heuristics, we will need to establish other signals that more closely track threateningness. Candidates range from unambiguous (wearing a ski mask in warm weather, aiming a visible weapon, etc.) to subtle (following an individual too closely for too long, visible gang membership, etc.). The former may be insulated from the effects of bias; if so, we can fairly exonerate some mistaken defenders by using a public evidential norm constrained to only the unambiguous signals.

Rendering the more ambiguous signals free from bias might very well require increased familiarity with minority and sub-cultures, so that members of the majority culture can quickly and reliably distinguish between the likely-dangerous members of a culture and the other members (i.e., distinguish a kid who likes hip-hop from a gang member.) If so, then the obligation to replace perverse with permissible signals generates an obligation to develop at least this level of familiarity with the minority culture. Agents who mistakenly impose defensive force because they lack the ability to distinguish threats in this

[38] Lee (2013) recommends that jurors be instructed not to allow themselves to be affected by bias, and, if they are ever unsure, to consider the case with the races of the defendant and victim reversed. Richardson and Goff (2012) survey much of the data about racial bias that I have presented here, and conclude that (i) use of race as a 'suspicion heuristic' is widespread in the US, (ii) this unjustly increases Black agents' risk of suffering mistaken defense, and as a result (iii) the legal standard for reasonable belief must exclude beliefs based on the suspicion heuristic, treating sincere but biased beliefs as an imperfect (partially excusing) defense rather than as an exonerating justification. They also recommend requiring defenders to retreat if possible, in hopes that retreating will give defenders extra time and clarity, reducing errors.

[39] These results were found in Plant et al. (2005); Sim et al. (2013) report similar findings.

more fine-grained way have failed an obligation to be adequately sensitive to evidence, and so cannot be justified by the evidence they do have. Instead, their position is symmetric to that of an agent who culpably fails to develop the ability to distinguish between cellphones and guns in ordinary conditions: though their defensive action may be epistemically justified *given their actual beliefs and evidence*, they failed to act with due diligence and are responsible for errors stemming from this failure.[40]

C The Slow Fix

If we are unable to reform the evidential standards to insulate them from bias, it may well be that a strict fact-relative norm is the only fair regulative norm. This is a highly revisionist and probably socially costly recommendation. After all, strict norms hold agents who acted on very reasonable grounds (like Joubran) responsible for manslaughter, and encourage would-be defenders to hesitate, which increases their odds of suffering wrongful harm. Despite these costs, if reform fails, adopting a strict norm is our only morally permissible option.

This can be thought of as a slow fix: by increasing the cost of false positive errors it encourages defensive agents to be more cautious, in turn reducing their readiness to rely on error-producing heuristics. Changing the costs in this way is unlikely to alter defenders' *conscious* thought process; often the decision to engage in self-defense is made in an immediate rather than reflective way. Nevertheless, the data suggest that allowing agents a few more moments in making their determination greatly increases accuracy. Optimistically, making agents hesitate just a second longer could drastically reduce their propensity to make racialized errors. Over time this should reduce the role that racialized cognitive biases play in determining which agents are subjected to mistaken defensive harm, with the result that minorities are progressively less likely to suffer unjust harms. In the best case scenario, the cancellation of cognitive bias allows for the eventual introduction of permissible heuristics and the adoption of an evidential norm.

In the worst case scenario, there is no stably permissible heuristic, as various biases keep re-emerging. In that case, strict liability will eventually tend toward the most bias-free distribution of costs achievable, with agents' *ex-ante* chances of suffering unjust harm largely determined by their own risk aversion strategy and brute luck. It is possible, though unlikely, that the cognitive biases are so entrenched that they will be unresponsive even to the most drastic change in the cost of false-positive error. Strict liability under these conditions is in many ways inferior to retaining an evidential norm; members of the minority will still be disproportionately likely to suffer mistaken defensive harm, and now in addition mistaken defenders will be held accountable at least for manslaughter. Nevertheless this outcome is a significant *moral* improvement over unjustly maintaining an evidential norm in one important respect: in holding mistaken agents responsible for their mistakes, it treats members of the minority as moral equals with equally strong claims to rights against the imposition of harm, and recognizes the unjustified

[40] My central focus in making these prescriptions is the case of individual citizen self-defense. It's not obvious how these considerations affect the permissibility conditions of police use-of-force. We might think that, *qua* agents of the state, police ought to be held to a higher standard than the average citizen in use-of-force decisions. We might also think that since officers can expect a greater proportion of their interactions to be with unjust aggressors, and that they (to some extent) accept these higher risks in order to provide a good to the community at large, we should be more tolerant of their false-positive errors. An adequate treatment of these issues demands a full-length paper of its own, so I will simply set them aside here. My thanks to Suzy Killmister for insightful discussion on this point.

imposition of harm as a violation of those rights.

VI Conclusion

There is some leeway in choosing a regulative norm for defense, but it must adjudicate fairly between the security interests of would-be defenders and the others' freedom to live without fear of suffering mistaken defensive harm. We have good reasons to prefer a norm that permits some reasonable mistakes, exonerating mistaken defenders when the apparent aggressor is responsible for the appearance of a threat. I outlined three constraints to ensure that the chosen norm is substantively just: (1) that it in-principle allow all agents fair opportunity to avoid being subject to defensive harm, (2) that it enable agents to predict and avoid behaviors that will make them subject to defensive harm, and (3) that it be applied equitably.

This involves a subtle re-framing of questions about mistaken defense. When deciding whether to fully exonerate an error, we are not merely making a ruling on whether the defender had intelligible reasons for his action; we are also making a pronouncement on whether his victim was wronged by the error or responsible for it. Focusing on the structural aspect of the problem highlights the fact that our choice of regulative norm is answerable not only to prospective defenders like Joubran, but also to those who, like Wong, are at risk of being mistaken for aggressors. I've argued that evidentialist regulatory norms are permissible only if they are not undermined by widespread bias, but we have compelling data showing that our actual circumstances include precisely this kind of pervasive bias. Unrestricted evidential norms adopted against this background are unjust in a very particular way: they fail to afford a subgroup of agents adequate opportunity to avoid appearing threatening. Adding insult to injury, a compromised evidential norm will fail even to recognize that bias-driven mistakes violate the rights of the agents who were mistakenly killed or assualted. Such a norm fails to respect them as moral equals.

It is unjust to simply continue embracing an evidentialist norm that has been compromised by bias; we are obligated to attempt reform. We may take one of two paths. Ideally, we can try to render the evidential norm permissible by restricting its scope: eradicating perverse heuristics and replacing them where possible with signals that do not impose unreasonable costs on any group in the community. If all attempts at this sort of reform fail, however, we must acknowledge that we cannot find such a minimally acceptable signal and embrace a strict norm. The one thing we cannot do is continue on with a compromised norm, assuring ourselves that because we only exonerate 'reasonable' mistakes, we have done what justice requires.

References

Anderson, E. (2010). *The Imperative of Integration.* Princeton University Press.

Armour, J. (1994). Race isa loquitur: of reasonable racists, intelligent bayesians, and involuntary negrophobes. *Stanford Law Review*, 46(4):781–816.

Bolinger, R. J. (ms). The moral foundations of mistaken self-defense.

Correll, J., Wittenbrink, B., Park, B., and Judd, C. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83(6):1314–1329.

Correll, J., Wittenbrink, B., Park, B., Judd, C., and Goyle, A. (2011). Dangerous enough: Modeling racial bias with contextual threat cues. *Journal of Experimental Social Psychology*, 47:184–189.

Correll, J., Wittenbrink, B., Park, B., Judd, C., Sadler, M., and Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology*, 92(6):1006–1023.

Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, 34(4):590–598.

Eberhardt, J. L., Purdie, V. J., Goff, P. A., and Davies, P. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology*, 87(6):876–893.

Ferzan, K. K. (2012). Culpable aggression: The basis for moral liability to defensive killing. *Ohio State Journal of Criminal Law*, 9:699.

Fletcher, G. (1988). *A Crime of Self Defense: Bernhard Goetz and the Law on Trial*.

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

Greenwald, A., Oakes, M., and Hoffman, H. (2003). Targets of discrimination: Effects of race on responses to weapons holders. *Journal of Experimental Social Psychology*, 39:399–405.

Hunt, J. (2015). Race, ethnicity, and culture in jury decision making. *Annual Review of Law and Social Science*, 11:269–88.

Jeffers, C. (2013). Should black kids avoid wearing hoodies? In Yancy, G. and Jones, J., editors, *Pursuing Trayvon Martin: Historical contexts and Contemporary Manifestations of Racial Dynamics*, pages 129–140. Lexington Books.

Kahn, K. B. and Davies, P. (2010). Differentially dangerous? phenotypic racial stereotypicality increases implicit bias among ingroup and outgroup members. *Group Processes and Intergroup Relations*, 14(4):569–580.

Katz, J. (2015). No retrial for north carolina officer who killed unarmed man. *New York Times*.

Kennedy, R. (2015). Lifting as we climb. *Harper's Magazine*.

Klauer, K. C. and Voss, A. (2008). Effects of race on response latencies in the weapon identification task: A test of six models. *Personality and Social Psychology Bulletin*, 4(8):1124–1140.

Lait, M. and Glover, S. (2001). 2nd panel says police slaying was justified. *Los Angeles Times*.

Larson, J. (2004). Behind the death of timothy thomas: Shooting of 19-year old brings to light pattern of ticketing that raises questions of racial profiling. *Dateline NBC*, http://msnbc.msn.com/id/4703574/.

Lazar, S. (2014). In dubious battle: Uncertainty and the ethics of killing. *(unpublished manuscript)*.

Lee, C. (2003). *Murder and the Reasonable Man: Passion and Fear in the Criminal Courtroom*. New York University Press.

Lee, C. (2013). Making race salient: Trayvon martin and implicit bias in a not yet post-racial society. *North Carolina Law Review*, 91:1555–1612.

Ma, D. S. and Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology*, 47:391–396.

McMahan, J. (2009). *Killing in War*. Oxford University Press.

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

PaynePayne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81:181–192.

Plant, E. A., Peruche, B. M., and Butz, D. A. (2005). Eliminating automatic racial bias: Making race non-diagnostic for responses to criminal suspects. *Journal of Experimental Social Psychology*, 41:141–156.

Rawls, J. (1999). *A Theory of Justice*. Belknap Press, revised edition.

Raz, J. (1986). *The Morality of Freedom*. Clarendon Press, Oxford.

Richardson, L. S. and Goff, P. A. (2012). Self-defense and the suspicion heuristic. *Iowa Law Review*, 98:293–336.

Ridley, G. (2012). Flint store clerk says shooting robber armed with plastic gun was right decision.

Roman, J. (2013). Race, justifiable homicide, and stand your ground laws: Analysis of fbi supplementary homicide report data. *Urban Institute*.

Ross, C. (2015). A multi-level bayesian analysis of racial bias in police shootings at the county-level in the united states, 2011-2014. *PLoS One*.

Sagar, H. A. and Schofield, J. W. (1980). Racial and behavioral cues in black and white children's perception of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, 39(4):590–598.

Scanlon, T. (2008). *Moral Dimensions*. Cambridge University Press.

Semuels, A. (2014). Detroit-area man gets 17 to 32 years for shooting visitor on porch. *Los Angeles Times*, http://www.latimes.com/nation/nationnow/la-na-nn-porch-killer-sentenced-20140903-story.html.

Sim, J., Correll, J., and Sadler, M. (2013). Understanding police and expert performance: When training attenuates (vs. exacerbates) stereotypic bias in the decision to shoot. *Personality and Social Psychology Bulletin*, 39(3):291–304.

Skyrms, B. (2010). *Signals*. Oxford University Press.

Sommers, S. and Ellsworth, P. (2000). Race in the courtroom: Perceptions of guilt and dispositional attributions. *Personality and Social Psychology Bulletin*, 26:1367–1379.

Sommers, S. and Ellsworth, P. (2001). White juror bias: An investigation of prejudice against black defendants in the american courtroom. *Psychology, Public Policy, and Law*, 7(1):201–229.

Steele, C. M. (2011). *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do*. W.W. Norton & Company.

Steiner, H. (1994). *An Essay on Rights*. Oxford: Blackwell.

Swaine, J. (2015). The counted: number of people killed by police this year reaches 500. *The Guardian*.

Swaine, J., Laughland, O., and Lartey, J. (2015). The counted: Black americans killed by police twice as likely to be unarmed as white people. *The Guardian*.

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

Uniacke, S. (1994). *Permissible Killing: The Self-defence justification of homicide*. Cambridge University Press.

Zipursky, B. (2015). Reasonableness in and out of negligence law. *University of Pennsylvania Law Review*, 163(2131).