


1-1-2008

Self-Defense and the Mistaken Racist

Stephen P. Garvey

Cornell Law School, spg3@cornell.edu

Follow this and additional works at: <http://scholarship.law.cornell.edu/facpub>

 Part of the [Criminal Law Commons](#), and the [Psychology and Psychiatry Commons](#)

Recommended Citation

Garvey, Stephen P., "Self-Defense and the Mistaken Racist" (2008). *Cornell Law Faculty Publications*. Paper 254.
<http://scholarship.law.cornell.edu/facpub/254>

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarship@Cornell Law: A Digital Repository. It has been accepted for inclusion in Cornell Law Faculty Publications by an authorized administrator of Scholarship@Cornell Law: A Digital Repository. For more information, please contact jmp8@cornell.edu.

SELF-DEFENSE AND THE MISTAKEN RACIST

Stephen P. Garvey*

How should the law of a liberal state respond when one person (D) kills another person (V), who is black, because D believes that V is about to kill him, but D would not have believed that V was about to kill him if V had been white? Should D be exonerated on grounds of self-defense? Some commentators argue that D's claim of self-defense should be rejected and that he should be convicted, either of murder or manslaughter, and punished accordingly.

I disagree. I argue that denying D's claim of self-defense would be at odds with the principle that an actor should be punished, not for possessing or choosing to possess racist or otherwise illiberal beliefs or desires, but only for choosing to cause (or risk causing) harm when the law does not permit him to make such a choice. Moreover, insofar as this principle can fairly be characterized as one to which a liberal state must adhere, then a liberal state should acknowledge D's claim of self-defense.

INTRODUCTION

Bernhard Goetz, a thirty-seven-year old white man with a “slight[] build,”¹ boarded a New York subway three days before Christmas in

*Professor of Law, Cornell Law School. Thanks to John Blume, Steven Clymer, Sherry Colb, Joshua Dressler, Sheri Johnson, Trevor Morrison, Christopher Seeds, Emily Sherwin, Joseph Wagner, participants at UCLA's Legal Theory Workshop, and the students in my Criminal Law Theory seminar for helpful comments on an earlier draft. James Rogers provided exemplary research assistance. Thanks also to the two anonymous reviewers for the *New Criminal Law Review*.

1. George P. Fletcher, *A Crime of Self-Defense: Bernhard Goetz and the Law on Trial* 1 (1988).

New Criminal Law Review, Vol. 11, Number 1, pps 119–171. ISSN 1933-4192, electronic ISSN 1933-4206. © 2008 by the Regents of the University of California. All rights reserved. Please direct all requests for permission to photocopy or reproduce article content through the University of California Press's Rights and Permissions website, <http://www.ucpressjournals.com/reprintInfo.asp>. DOI: 10.1525/ndr.2008.11.1.119.

December 1984.² Four “noisy and boisterous”³ young men,⁴ all of them black,⁵ were also onboard. Two of the young men approached Goetz.⁶ “Give me five dollars,”⁷ one of them said. When the request was repeated,⁸ Goetz opened fire with a concealed .38 caliber pistol loaded with five rounds.⁹ He fired four shots in rapid succession, wounding three of the victims.¹⁰ After pausing to survey the scene,¹¹ Goetz approached the fourth victim and said to him, “You seem to be all right, here’s another.”¹² He then fired the final round. The bullet severed the victim’s spinal cord, leaving him paralyzed.¹³ Most of the other passengers fled when the shooting started, but two women, one of whom was black,¹⁴ remained, “immobilized by

2. The statement of the facts that follows is based primarily on the principal opinion of the New York Court of Appeals, which was in turn based heavily on pretrial statements Goetz made to the police. *People v. Goetz*, 497 N.E.2d 41, 43–44 (N.Y. 1986). The Court of Appeals reversed a trial-court order dismissing certain counts of a second grand jury’s multiple-count indictment for attempted murder, assault, illegal possession of a firearm, and reckless endangerment. See *People v. Goetz*, 502 N.Y.S.2d 577 (Sup. Ct.), *aff’d*, 501 N.Y.S.2d 326 (N.Y. App. Div.), *rev’d*, 497 N.E.2d 41 (N.Y. 1986). Goetz was finally convicted on one count of criminal possession of a firearm in the third degree. He was sentenced to one year in prison, five years’ probation, and a \$5,000 fine. See *People v. Goetz*, 529 N.Y.S.2d 782 (App. Div.), *aff’d*, 532 N.E.2d 1273 (N.Y. 1988); see also Ronald Sullivan, *Goetz Is Given One-Year Term on Gun Charge*, *N.Y. Times*, Jan. 14, 1989, at B1.

For book-length accounts of the case, see Fletcher, *supra* note 1; Mark Lesly & Charles Shuttlesworth, *Subway Gunman: A Juror’s Account of the Bernhard Goetz Trial* (1988); and Lillian B. Rubin, *Quiet Rage: Bernie Goetz in a Time of Madness* (1986).

3. See Fletcher, *supra* note 1, at 1.

4. Two of the victims (James Ramseur and Barry Allen) were eighteen. The other two (Darrell Cabey and Troy Canty) were nineteen. *Id.* at 2–3.

5. *Id.* at 1.

6. *Goetz*, 497 N.E.2d at 43 (“Canty approached Goetz, possibly with Allen beside him.”).

7. *Id.*

8. *Id.* at 44.

9. *Id.* at 43.

10. *Id.*

11. *Id.* at 44. But cf. Fletcher, *supra* note 1, at 171 (noting that “eight witnesses testified that they did not hear a pause”).

12. *Goetz*, 497 N.E.2d at 44.

13. *Id.* The fourth victim was Darell Cabey. Cabey later filed a civil suit against Goetz and won a \$43 million judgment in 1996. Civil Complaint Against “Subway Vigilante” Bernhard Goetz Filed ‘85’ and Tried ‘96’, <http://www.lectlaw.com/files/cas91.htm> (last visited Jan. 4, 2008).

14. Fletcher, *supra* note 1, at 5.

fear.”¹⁵ Goetz “sa[id] some soothing words”¹⁶ to them and jumped off the train. He later turned himself in.¹⁷

Charged with attempted second-degree murder,¹⁸ Goetz claimed that he acted in self-defense, according to which an actor is, generally speaking, permitted to use deadly force against an aggressor if—but only if—he *reasonably* believed that the use of deadly force was necessary to avoid death or grievous bodily harm to himself.¹⁹ In other words, what the defendant believed were the facts at the time of the crime is not what matters, or more

15. *Id.* at 2.

16. *Id.*

17. Goetz, 497 N.E.2d at 44.

18. See *id.* at 43. The jury acquitted Goetz on all charges (including the attempted murder charges), with the single exception of the firearms possession count. Although “the defense [had] never seriously challenged whether, as a matter of fact, Goetz intended to cause death by shooting the four youths,” Fletcher, *supra* note 1, at 185, the jury nonetheless believed that he lacked the intent needed to convict on attempted murder, see, e.g., *id.* at 186–88; Lesly, *supra* note 2, at 279–82, and so never reached the question of self-defense as an affirmative defense. The jurors believed that Goetz lacked the requisite intent because they “incorporated Goetz’s purpose of defending himself into their analysis of his intention [to kill].” Fletcher, *supra* note 1, at 187. For the jury, Goetz’s intent was to defend himself, and in order to do that, he intended to pull the trigger on the gun. But he did not intend to cause death. Causing death was merely a foreseeable consequence of pulling the trigger. Although this way of thinking about self-defense is generally at odds with contemporary thinking, see, e.g., *id.* at 186, it nonetheless has a long history rooted in the Catholic doctrine of double effect. See, e.g., Fiona Leverick, *Killing in Self-Defence* 53–54 (2006) (describing “appeal[s] to the doctrine of double effect” as an “explan[ation for] the justification of killing in self-defence” but concluding that the attempted explanation is ultimately unpersuasive). The underlying philosophical question at issue here deals with how one should go about individuating the objects of intention. See, e.g., Michael Moore, *Placing Blame: A General Theory of the Criminal Law* 469 (1997) (arguing in favor of a “very fine-grained theory” of individuation).

19. See, e.g., Joshua Dressler, *Understanding Criminal Law* § 18.01[E], at 239 (4th ed. 2006) (“A defendant is justified in killing a supposed aggressor if the defendant’s belief in this regard is objectively reasonable.”); Wayne R. LaFare, *Criminal Law* § 5.7(c), at 493–95 (3d ed. 2000) (“[S]elf-defense generally require[s] that the defendant’s belief in the necessity of using [deadly] force to prevent harm to himself be a reasonable one. . . .”); Leverick, *supra* note 18, at 161 (“The majority of common law jurisdictions take the approach that only a reasonable mistake about the existence of an attack should be permitted to ground an acquittal.”); Rollin M. Perkins & Ronald N. Boyce, *Criminal Law* 1127 (3d ed. 1982) (“[D]eadly force is authorized to defend against deadly force if this reasonably seems necessary to avoid death or great bodily injury.”).

At the time Goetz was tried, New York law permitted (and continues to permit) an actor to use deadly physical force against another person when he reasonably believes that the

precisely, not all that matters, in the law of self-defense.²⁰ He must believe that he needed to use deadly force in order to avoid being the victim of deadly force, *and* it must have been reasonable for him to have so believed. The facts themselves, however, do not matter.²¹ An actor who kills because he reasonably believes that he is about to be killed is entitled to an acquittal on grounds of self-defense even if it later turns out that he was wrong. Again, what matters is what the actor reasonably believed were the facts, not the facts themselves. Call this the *reasonable-belief rule*.²²

“other person is using or about to use deadly physical force,” N.Y. Penal Law § 35.15(2)(a) (McKinney 2007), where deadly physical force means “physical force which, under the circumstances in which it is used, is readily capable of causing death or serious physical injury.” Id. § 10.00(11). It also permitted (and continues to permit) an actor to use deadly physical force against another person when he reasonably believes that the other person is “committing or attempting to commit a . . . robbery,” id. § 35.15(2)(b), where robbery is defined as larceny in which the actor “uses or threatens the immediate use of physical force upon another person,” id. § 160.00, whether or not that physical force rises to the level of deadly physical force. See also 2 Paul H. Robinson, *Criminal Law Defenses* § 131(d), at 83 (“Some states expressly authorize the use of deadly defensive force in response to certain enumerated offenses.”).

20. An actor who kills because he unreasonably believed that he was about to be killed may in some jurisdictions be entitled to a partial defense usually described as “imperfect self-defense.” See Dressler, *supra* note 19, § 18.03, at 249.

21. The facts may matter as to whether the defense is characterized as an excuse or as a justification. See id. § 18.04[A]–[B], at 250–51.

22. According to a recent annotation, the reasonable-belief rule is the majority rule in the United States. See John F. Wagner, Jr., Annotation, Standard for Determination of Reasonableness of Criminal Defendant’s Belief, for Purposes of Self-Defense Claim, That Physical Force Is Necessary—Modern Cases, 73 A.L.R.4th 993, 996 (1989 & Supp. 2006) (“[M]ost states having penal codes have . . . opted for a ‘reasonable belief’ rule . . .”). However, according to at least one writer, it was not always so. See Richard Singer, *The Resurgence of Men Rea: II—Honest But Unreasonable Mistake of Fact in Self-Defense*, 28 B.C. L. Rev. 459, 470–90 (1987) (arguing that the reasonable-belief rule was not incorporated into American law until the mid-nineteenth century).

Under English law, an actor who honestly but unreasonably believed that he needed to use force to protect himself from force is deemed to lack the intent to inflict “unlawful” force. Consequently, an honest belief in the need to use deadly force is sufficient to preclude conviction. See *Regina v. Williams*, [1984] 78 Crim. App. 276, 281 (“In a case of self-defence . . . if the jury came to the conclusion that the defendant believed . . . that force was necessary to protect himself . . . , then the prosecution have not proved their case.”); *Beckford v. Regina*, [1987] 85 Crim. App. 378, 385 (“[A] genuine belief in facts which if true would justify self-defence [must] be a defence to a crime of personal violence because the belief negates the intent to act unlawfully.”). For

The *Goetz* case was at the time a “cause célèbre,”²³ and it has since passed into the canon of the criminal law, at least in the United States.²⁴ Few American lawyers will have left law school without having encountered it. Nonetheless, my present goal is neither to analyze nor comment on the actual case itself, which is rich in detail as well as controversy. Instead, my goal is to better understand the law of self-defense, and in particular its insistence that an actor is entitled to claim self-defense if and only if his belief that he needed to use deadly force in order to avoid death or grievous bodily injury was reasonable.

In order to accomplish that goal without needless distraction related to the real case of Bernhard Goetz, or any real case for that matter, I ask you to imagine another defendant, who I will call Goetz*. Goetz*, unlike the real

commentary on English law, see Andrew Ashworth, *Principles of Criminal Law* § 6.6, at 235 (4th ed. 2003) (“[A] putative defence will succeed whenever D raises a reasonable doubt that he actually held the mistaken belief, no matter how outlandish that belief may have been.”); A.P. Simester & G.R. Sullivan, *Criminal Law: Theory and Doctrine* § 17.1(iii), at 550 (2d ed. 2003) (“[A] person who believed force was necessary to protect another from violence would lack an intent to inflict unlawful force.”); William Wilson, *Criminal Law: Doctrine and Theory* § 9.10(B), at 253 (2d ed. 2003) (“The . . . requirement that the mistake made be a reasonable one was abandoned [in *Williams*].”); Andrew Simester, *Mistakes in Defence*, 12 *Oxford J. Legal Stud.* 295, 295 (1992) (arguing that the “reasoning in [*Williams* and *Beckford*] is unsound and has unfortunate implications for the criminal law in general”). Cf. George Fletcher, *Mistake in the Model Penal Code: A False False Problem*, 19 *Rutgers L.J.* 649, 652 (1988) (noting that “[i]f every relevant factual issue were intrinsic to the required intent, any mistake would be a good defense”).

23. Sanford H. Kadish et al., *Criminal Law and Its Processes: Cases and Materials* 743 (8th ed. 2007). But see Franklin E. Zimring, *Hardly the Trial of the Century*, 87 *Mich. L. Rev.* 1307, 1309 (1989) (book review) (“[W]hat is there about [the *Goetz* case] that justifies its landmark status in public discussions of crime and criminal justice? Perhaps there is less than we might suppose.”).

24. The case is reproduced in several criminal-law casebooks. See Richard J. Bonnie et al., *Criminal Law* 419 (2d ed. 2004); Ronald N. Boyce et al., *Criminal Law and Procedure: Cases and Materials* 912 (10th ed. 2007); Joseph G. Cook & Paul Marcus, *Criminal Law* 691 (4th ed. 1999); Joshua Dressler, *Cases and Materials on Criminal Law* 504 (4th ed. 2007); Markus D. Dubber & Mark G. Kelman, *American Criminal Law: Cases, Statutes, and Comments* 542 (2005); Kadish et al., *supra* note 23, at 739; John Kaplan et al., *Criminal Law: Cases and Materials* 521 (5th ed. 2004); Wayne R. LaFare, *Modern Criminal Law: Cases, Comments and Questions* 510 (4th ed. 2006); Cynthia Lee & Angela Harris, *Criminal Law: Cases and Materials* 727 (2005); Andre A. Moenssens et al., *Criminal Law: Cases and Comments* 518 (7th ed. 2003); Paul H. Robinson, *Criminal Law: Case Studies and Controversies* 559 (2005).

Goetz, fired only one shot from his .38, not all five,²⁵ and he hit only one of the young men, not all four. The other three fled unharmed. Moreover, in order to avoid unnecessary doctrinal complications arising from the law of attempts,²⁶ I also ask you to imagine that Goetz*'s first and only shot killed its intended victim, rather than simply wounding him. Finally, whatever the real Goetz believed, I ask you to assume that Goetz* did indeed believe that he was about to be killed, and that he needed to use deadly force in order to avoid being killed.²⁷ I will hereafter refer to this belief—I am about to be killed, and deadly force is necessary to avoid being killed—as the belief that *p*.

The principal question the Goetz* case raises is this: Who *is* the reasonable person? According to the standard analysis, we need to answer that question in order to know whether or not it was reasonable for Goetz* to have believed that *p*, and therefore whether or not he should have been acquitted on grounds of self-defense or convicted of murder. Moreover, according to the standard analysis, that question in turn gives rise to another: Which characteristics of the real defendant should be imputed to the reasonable person?²⁸ One answer is that *all* of them should be imputed. But that answer

25. According to Fletcher, “[i]f Goetz really did pause after the fourth shot, physically approach [the final victim], and say, ‘You seem to [be doing] all right; here’s another,’ it would be almost impossible to construe this shot as a reasonable act of self-defense.” Fletcher, *supra* note 1, at 170. But see *id.* at 171 (noting that the testimony of other witnesses suggested no such pause); Singer, *supra* note 22, at 516 (“[E]ven Goetz’s fifth shot could be found by a jury to have emanated from a swirl of anxiety and loss of control which continued far after the last shot.”).

26. See, e.g., Dressler, *supra* note 19, § 27.05, at 417–22 (discussing the mens rea of attempts).

27. Although the discussion hereafter proceeds on the facts of Goetz*, and not on the facts of the real Goetz case, I nonetheless suspect, though I could of course be wrong, that those who would have been disinclined to acquit the real Goetz will also be disinclined to acquit Goetz*. I have this suspicion because I doubt that the features I strip away from the real case are really what matters to most people who believe that the real Goetz did not act in self-defense. What matters to them is the fact that Goetz believed that he was about to be attacked because he was a racist, and would not have believed that he was about to be attacked had he not been a racist. But I will be assuming that those facts are true of Goetz* too.

28. See, e.g., Dressler, *supra* note 19, § 18.05, at 253 (“The crux of the issue, at least as courts see the matter, is: . . . [T]o what extent should courts permit juries, as factfinders, to incorporate the defendant’s own characteristics or life experiences in the ‘reasonable person’ standard?”); Cynthia Lee, *Murder and the Reasonable Man* 209 (2003) (“Since most jurisdictions utilize a hybrid subjectivized-objective standard, a critical question is which of the defendant’s characteristics are or should be incorporated into the Reasonable Person standard?”).

Paul Robinson and Michael Cahill have noted that “criminal-law theorists have not yet been able to articulate a comprehensive principle that defines what should and should not

is a nonstarter: imputing all of the defendant's characteristics to the reasonable person would mean that the reasonable person just *is* the defendant, such that any belief the defendant possessed would be "reasonable," just because he possessed it.²⁹ Beyond that, things get less clear. Which characteristics are in, and which are out? Thankfully, in the context of a case like *Goetz**, the discussion usually focuses on the most salient of the defendant's characteristics: his putative racism. The principal question can therefore provocatively be put thus: Is the reasonable person a racist?

That is the standard analysis. I want to offer an alternative, which proceeds as follows. The choices we make at any moment in time, like the

be allowed to individualize the reasonable-person standard" and that this question is "perhaps the greatest challenge to the present and coming generation of theorists . . ." Paul H. Robinson & Michael T. Cahill, *Law Without Justice: Why Criminal Law Doesn't Give People What They Deserve* 51 (2006). One possible explanation for this state of affairs may be, as Larry Alexander famously argued, that any answer to the question as formulated is bound to be "morally arbitrary." Larry Alexander, *Reconsidering the Relationship among Voluntary Acts, Strict Liability, and Negligence in Criminal Law*, 7 Soc. Phil. & Pol'y, Spring 1990, at 84, 99. If so, then Peter Westen is doubtless correct when he says that the question is the wrong one to be asking. See Peter Westen, *Individualizing the Reasonable Person in Criminal Law*, *Crim. L. & Phil.* (forthcoming 2008) (manuscript on file with author) (arguing that the right question is: "What would a person, who otherwise possessed every trait of the actor but fully respected the interests that the statute at hand seeks to protect, have [believed] on the occasion at issue?"); see also R.A. Duff, *Choice, Character, and Criminal Liability*, 12 *Law & Phil.* 345, 359 (1993) ("[W]e should give the reasonable person any of this defendant's actual characteristics . . . other than [those] which involve or reveal a lack of proper regard for the law and its values. . ."). I agree with Westen that we are asking the wrong question, but my proposed replacement is different.

29. See, e.g., George P. Fletcher, *Rethinking Criminal Law* § 6.8, at 513 (1978) ("If the reasonable person were defined to be just like the defendant in every respect, he would arguably [believe and] do exactly what the defendant [believed and] did under the circumstances. Thus the standard of judgment collapses into a description of the particular defendant."); Robinson & Cahill, *supra* note 28, at 50 ("[A] complete individualization of the objective standard . . . would produce a purely subjective standard.").

One might argue that a fully subjective standard does not in fact eliminate the reasonable-belief requirement altogether. The idea would be that under a fully subjective standard an actor's belief that *p* is a reasonable belief if the actor believes that *p* and at the same time believes that it is reasonable to believe that *p*. Conversely, an actor's belief that *p* is an unreasonable belief if the actor believes that *p* but at the same time believes that it is unreasonable to believe that *p*. An actor in this latter epistemic state can be described as epistemically akratic. He believes that *p* at the same time that he believes all things considered that he should not believe that *p*, just as a practically akratic actor performs an act at the same time that he believes all things considered that he should not perform that act. A debate exists as to whether this epistemic state is conceptually impossible or merely irrational.

choice to kill, depend on the beliefs we possess at that moment. Our choices, moreover, are up to us. We can choose or not as we see fit. In contrast, the beliefs we possess at any moment are *not* up to us. We can choose to act or not act on our beliefs, but we cannot choose our beliefs. Thus, although Goetz* did not choose to believe that *p*, he could have chosen not to act on that belief. But the belief that *p*—I am about to be killed, and deadly force is necessary to avoid being killed—is one that only a saint or a fool would ignore. An actor who believes that he is about to be killed could remain passive, but why should he? What good reason would he have to do nothing? If no such reason is forthcoming, then I would argue that self-defense should, all else being equal, be available to any actor who killed because and only because he believed that doing so was necessary to avoid being killed or seriously injured. Such an actor chooses to kill, but from where he stands, he had no other choice, and where else could he have stood?

Let me pause here in order to anticipate and attempt to disarm one likely objection. A law of self-defense based exclusively on a defendant's honest belief that *p* would, so the objection goes, provide far too little protection to those innocents who find themselves on the receiving end of the mistaken defendant's deadly force. But is that true? It is hard to see how it could be. The objection presupposes that an actor who is aware of the reasonable-belief rule will not use deadly force if and when he believes that his belief that *p* is unreasonable. That supposition in turn presupposes that an actor can believe that *p* and at the same time believe that believing that *p* is unreasonable (and presumably that he should therefore not believe that *p*).³⁰ If that supposition is false, if an actor cannot believe that *p* and at the same time believe that he should not believe that *p*, then the reasonable-belief

Compare Jonathan E. Adler, *Akratic Believing?*, 110 *Phil. Stud.* 1, 21 (2002) (“[T]he first-personal thought corresponding to the admission of akratic belief would be not merely irrational, but incoherent.”); David Owens, *Epistemic Akrasia* 85 *Monist* 381, 395 (2002) (“[E]pistemic akrasia . . . is impossible.”), with John Heil, *Doxastic Incontinence*, 93 *Mind* 56, 65 (1984) (“Doxastic incontinence is reprehensible, not because it holds out an unattainable goal, but because it is at odds with what we take to be the aims of rational doxastic agents.”); Alfred R. Mele, *Incontinent Believing*, 36 *Phil. Q.* 212, 217 (1986) (arguing that “full-blown incontinent believing” is “possible”).

30. An actor who finds himself in simultaneous possession of such beliefs might be said to be suffering from “epistemic akrasia.” For more on this idea, see *supra* note 29.

rule is inert, exerting no influence on the behavior of an actor who believes that p .³¹ Consequently, a law of self-defense based on the reasonable-belief rule would provide no greater protection to innocent victims than would one based on honest belief alone.

Again, I claim that self-defense should be available, all else being equal, to any actor who believes that p , including one like Goetz*. Nonetheless, I make no argument here for abolishing the reasonable-belief rule, for all else may *not* be equal. For example, imagine two actors, each of whom believes that p . Imagine further that the second actor, but not the first, believes that p only because he culpably violated some other obligation with respect to which the state can legitimately demand compliance. If so, then the law might elect to deny him, in whole or in part, the claim of self-defense to which he would otherwise have been entitled. Moreover, a belief formed as a result of such a violation might fairly be described as “unreasonable.”³² In other words, I suggest (in part I) that the reasonable-belief rule should be portrayed as a *forfeiture rule*: an actor who unreasonably believes that p forfeits, in whole or in part, the claim of self-defense upon which he would otherwise have been permitted to stand.

According to this line of thought, if Goetz* killed because and only because he believed that he was about to be killed, and if the law nonetheless convicts him of murder, it does so because it holds him to have forfeited his claim to self-defense, and it holds him to have forfeited that claim because the only reason he believed that p was because he culpably violated some other obligation with respect to which the law can legitimately hold him to account. As such, whether Goetz* is guilty of murder, or whether he should be acquitted on grounds of self-defense, depends on whether any such obligation exists; and if so, whether that obligation is one with respect to which the state can legitimately demand compliance; and if so, whether Goetz* culpably violated it.

The purported unreasonableness of the real Goetz's belief that p (assuming that he did indeed believe that p) is usually associated with the

31. See, e.g., Dressler, *supra* note 19, § 18.04[A], at 251 (“One who is threatened with immediate death is not deterrable by the threat of criminal sanction. Therefore, his punishment is inefficacious.”).

32. Conversely, a belief formed in the absence of any such violation might fairly be described as “reasonable.”

idea that Goetz was a racist.³³ Accordingly, I assume for present purposes that Goetz* is a racist as well, and that he would not have believed that *p* had he not been a racist. Consequently, if the victim had been white, Goetz* would not have believed that he was about to be killed.³⁴ With those assumptions in hand, at least three different theories can be offered to explain why Goetz*'s belief that *p* was unreasonable, each of which identifies an obligation that Goetz* is assumed to have breached and upon which the forfeiture of his self-defense claim is based. The pivotal question is whether the respective obligation upon which each of these theories rests is one the state can legitimately impose on its citizens.

33. Some evidence suggested the contrary. For example, according to an article written in *New York Magazine* soon after the first grand jury declined to indict the real Goetz on attempted-murder charges, Goetz's neighbor, Myra Friedman, wrote:

The other troubles of 14th Street[on which Goetz lived,] remained. People in the building who had always considered themselves to be liberals began expressing some surprising sentiments. Bernie was one of these people. At a community meeting, I heard him say, "The only way we're going to clean up this street is to get rid of the spics and niggers." I was shocked to hear a man who I knew to have close black and Hispanic friends talk this way, and I said, "I'm getting out of here." Later, somebody close to Bernie for many years suggested that he used an occasional racial epithet just to shock.

Myra Friedman, *My Neighbor Bernie Goetz*, N.Y. Mag., Feb. 18, 1985, at 34, 35. George Fletcher, who observed the proceedings against Goetz firsthand, concluded that "[w]e have to accept the implication that at the time of [Goetz's] confession, at least, racial consciousness and animosity did not weigh heavily in Goetz's mind." Fletcher, *supra* note 1, at 205.

34. The assumption that race can make the difference between forming the belief that *p* and not forming that belief is consistent with empirical studies showing that actors are more apt to perceive a threat when, all else being equal, the putative assailant is black than when he is white. See, e.g., Joshua Correll et al., *The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals*, 83 J. Personality & Soc. Psychol. 1314, 1325 (2002) (noting that participants in a study "fired on an armed target more quickly when he was African American than when he was White"); Charles M. Judd et al., *Automatic Stereotype vs. Automatic Prejudice: Sorting Out the Possibilities in the Payne (2001) Weapon Paradigm*, 40 J. Experimental Soc. Psychol. 75, 80 (2004) ("Black faces seem to facilitate weapon identification compared to White faces."); B. Keith Payne, *Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon*, 81 J. Personality & Soc. Psychol. 181, 182 (2001) ("[W]hen forced to respond rapidly, racial cues may cause perceivers to make stereotype-consistent errors."); B. Keith Payne et al., *Best Laid Plans: Effects of Goals on Accessibility Bias and Cognitive Control in Race-based Misperceptions of Weapons*, 38 J. Experimental Soc. Psychol. 384, 394 (2002). The participants in one of the studies reported in Correll et al. "consist[ed] of both Whites and African Americans." Correll et al., *supra*, at 1328. Correll and his co-authors concluded that their "studies . . . suggest that Shooter Bias is present

The first theory (discussed in part II) traces the unreasonableness of Goetz*'s belief that *p* to his character. According to this *character theory*, Goetz*'s belief that *p* was unreasonable because he was a racist (and should not have been), and his racism in turn caused him to form the belief that *p*. This theory links the unreasonableness of Goetz*'s belief that *p* to the fact that he violated an obligation not to possess a racist character. As a result of that violation, he forfeits his claim of self-defense.

The second and third theories (discussed in part III) trace the unreasonableness of Goetz*'s belief that *p* to a choice he made. According to the *belief-choice theory*, Goetz*'s belief that *p* was unreasonable because he chose to believe that *p* and that choice was based on racism. This theory links the unreasonableness of Goetz*'s belief that *p* to the fact that he violated an obligation not to choose to believe that *p*, when that choice is based on racism. As a result of that violation, he forfeits his claim of self-defense.

According to the *character-choice theory*, Goetz*'s belief that *p* was unreasonable, not simply because he was a racist, nor because he chose to believe that *p*, but rather because he chose to be or remain a racist. This theory links the unreasonableness of Goetz*'s belief that *p* to the fact that he violated an obligation not to choose to be or remain a racist, and his being a racist in turn caused him to form the belief that *p*. As a result of that violation, he forfeits his claim of self-defense.

I argue that while some states might be able to embrace one or more of these theories, a liberal state can embrace none of them. In one way or another each is inconsistent with the principle that an actor should be punished, not for possessing or choosing to possess racist or otherwise illiberal beliefs or desires, but only for choosing to cause (or risk causing) harm when the law does not permit him to make such a choice. Consequently, insofar as this principle is one to which a liberal state owes allegiance, and insofar as our criminal law aspires to be a liberal criminal law, Goetz* should be acquitted on grounds of self-defense, assuming once again that when he pulled the trigger he honestly believed that he was about to be killed. A liberal state should not, without more, punish a citizen for choosing to kill when and because he honestly believes he is about to be killed. If acquitting Goetz* on grounds of self-defense is thought to

among White college students . . . and among a community sample that consists of both Whites and African-Americans. . . ." Id. The other studies cited above included only non-black participants.

be the wrong result, a liberal state nonetheless has little choice but to accept it, unless it wishes to abandon the above-mentioned principle, or until some other theory of unreasonableness consistent with that principle can be identified and defended.³⁵

I. THE REASONABLE-BELIEF RULE

As a general proposition an actor who uses deadly force against another should be acquitted on grounds of self-defense if he reasonably believed that the use of such force was necessary to prevent his death or serious bodily injury.³⁶ In some jurisdictions the actor must also have reasonably

35. I make no claim that the theories examined here exhaust all the possibilities. What I mainly hope to accomplish is to place the argumentative burden of proof on those who believe that Goetz* can be punished consistent with the principle mentioned in the text.

36. The Model Penal Code requires that the actor's use of force be "immediately necessary." Model Penal Code § 3.04(1) (Proposed Official Draft 1962). No jurisdiction that I know of specifies how probable an actor must reasonably believe a lethal attack to be before he is permitted to use deadly force to preempt it. In other words, no statute defining self-defense says, for example, that an actor must reasonably believe that the probability of death or serious bodily injury is 75 percent before the actor can respond with deadly force. Likewise, no jurisdiction specifies how confident an actor must be in his belief that death or serious bodily injury is imminent before he is permitted to use deadly force. For present purposes, I will assume that an actor must have whatever measure of confidence is needed in order to say that his cognitive attitude toward *p* qualifies as a belief, and not merely a suspicion. If an actor merely suspects that *p*, then he probably should not be permitted to use deadly force. See Boaz Sangero, *Self-Defence in Criminal Law* 287–88 (2006) ("A situation could exist in which the actor is not sure whether he is being attacked or not. In such a situation there is good reason to require that he explore and verify the situation prior to using defensive force.").

Mark Kelman and Jody Armour nonetheless propose that an actor should be entitled to self-defense if and only if the error costs associated with a false positive (i.e., believing one is about to be attacked when one is not about to be attacked) are less than those associated with a false negative (i.e., believing one is not about to be attacked when one is about to be attacked). See Jody D. Armour, *Race Ipsa Loquitur*: Of Reasonable Racists, Intelligent Baysians, and Involuntary Negrophobes, 46 *Stan. L. Rev.* 781, 794–95 (1994); Mark Kelman, Reasonable Evidence of Reasonableness, 17 *Critical Inquiry* 798, 815–16 (1991). For example, suppose at the moment he pulled the trigger that Goetz* believed that the probability of him being killed (unless he killed first) was 75 percent. According to the Kelman-Armour thesis, Goetz* should not, despite his belief, be acquitted on grounds of self-defense if a jury decides that the false-negative error costs associated with requiring him to wait are less than the false-positive error costs associated with permitting him to kill. Moreover, while the false-negative error costs are limited more or less to Goetz*'s death, the false-positive error costs are not limited to the death of the innocent victim. Inasmuch as Goetz* "selected [his victims] on the basis of their race," Kelman, *supra*, at 815, those costs also include the stigmatization of

believed that his assailant's use of such force was imminent.³⁷ If an actor's beliefs regarding the elements of the defense were indeed reasonable,³⁸ he remains entitled to the defense even if it turns out later that he was wrong to believe that he was about to be attacked.

young black men and their consequent exclusion from full participation in public life. See *id.* at 816. Thus, despite his belief that the chance of him being killed unless he killed first was 75 percent, the law should demand that Goetz* wait until he believed that the chance was even higher, though how much higher is unclear.

This theory may be an attractive proposal for reforming self-defense law. It may even be an accurate description of how jurors actually go about deciding cases of self-defense. But it is not an accurate statement of the law of self-defense. The law of self-defense (generally speaking) permits an actor to use deadly force when he reasonably believes that the use of such force is necessary to prevent him from being killed or seriously injured. In contrast, the Kelman-Armour proposal would permit an actor to use deadly force if and only if, given the actor's belief that the probability that he will be killed or seriously injured (unless he kills first) is ϕ , the false-positive error costs of killing are less than the false-negative error costs of waiting. This proposal asks the jury to assess not the reasonableness of the actor's belief, but the reasonableness of his action in light of his beliefs. See *id.* at 800 (arguing that "although the stated norm in self-defense cases [i.e., the law of self-defense] makes reference only to the reasonableness of the defendant's factual perceptions, we in fact also expect the jury to judge the reasonableness of his decision to use deadly force, and that two defendants facing an equal chance of grievous bodily harm or death may not and should not always be judged to be acting equally reasonably in doing so"). Moreover, existing doctrine is already designed to make sure that an actor's use of deadly force is reasonable, at least in the sense that it is proportional. An actor can only use deadly force to defend against deadly force. He cannot use deadly force against nondeadly force, even when the use of deadly force is necessary to avoid nondeadly injury.

37. See, e.g., Dressler, *supra* note 19, § 18.02[D][1], at 246–48. One can imagine cases in which an actor reasonably believes that he is facing imminent death or serious bodily injury but nonetheless unreasonably believes that the use of deadly force is necessary to avoid such death or injury. If some measure of force less than that of deadly force would suffice to avoid an imminent threat of death or serious injury, then the actor is not permitted to resort to deadly force. He is only permitted to use the lesser force needed to avoid the threat. See, e.g., *id.*, § 18.02[C], at 238.

For arguments in favor of eliminating the imminence requirement from the law of self-defense, see 2 Robinson, *supra* note 19, § 131(c)(1), at 78 ("The proper inquiry is not the immediacy of the threat but the immediacy of the response necessary in defense."); Richard Rosen, On Self-Defense, Imminence, and Women Who Kill Their Batterers, 71 N.C.L. Rev. 371, 380 (1993) ("Because imminence serves only to further the necessity principle, if there is a conflict between imminence and necessity, necessity must prevail."). But see Kimberly Kessler Ferzan, Defending Imminence: From Battered Women to Iraq, 46 Ariz. L. Rev. 213, 217 (2004) (defending the imminence requirement on the grounds that "[i]mminence serves as the *actus reus* for aggression, separating those threats that we may properly defend against from mere inchoate and potential threats").

38. The Model Penal Code's self-defense provision does not speak in terms of "reasonable" or "unreasonable" beliefs. See Paul H. Robinson, Criminal Law § 4.4, at 263–64 (1997)

Most everyone agrees that an actor who uses deadly force when he reasonably and *correctly* believed that *p* is properly characterized as having been justified in using such force.³⁹ At the very least, the law permits him to use such force. In contrast, how best to characterize an actor who uses deadly force when he reasonably but *incorrectly* believed that *p* is a matter

(describing differences between common law and MPC approaches to mistakes generally). Instead, the code's self-defense provision says that an actor is permitted to use deadly force if he believes—reasonably or otherwise—that the use of such force is “necessary to protect himself against death, serious bodily injury, kidnapping or sexual intercourse compelled by force or threat. . . .” Model Penal Code § 3.04(2)(b) (Proposed Official Draft 1962). Standing alone, this provision would mean that an actor is entitled to an acquittal on grounds of self-defense if he honestly believed that the use of deadly force was necessary to protect himself against the itemized harms, no matter how unreasonable that belief might be.

But another section of the code qualifies this provision, such that if an actor is “reckless or negligent in having such belief or in acquiring or failing to acquire any knowledge or belief which is material to the justifiability of his use of force, the justification afforded by [the self-defense provision] is unavailable in a prosecution for an offense for which recklessness or negligence, as the case may be, suffices to establish culpability.” *Id.* § 3.09(2). This provision is usually understood to mean that an actor who recklessly believes that the use of deadly force is immediately necessary can, for example, raise the defense with respect to a charge of murder (for which recklessness does not suffice for liability), but not with respect to a charge of manslaughter (for which recklessness does suffice); likewise, an actor who negligently believes that the use of deadly force is immediately necessary can raise the defense with respect to a charge of murder or manslaughter (for which negligence does not suffice), but not with respect to a charge of negligent homicide (for which negligence does suffice). The code's approach has the virtue of trying to align the culpability associated with an actor's mistaken belief in the need to use deadly force with the offense for which he is ultimately held liable. A reckless mistake gets you reckless homicide; a negligent mistake gets you negligent homicide. Nonetheless, one problem with this approach (among others) is its reliance on the idea of a “reckless belief.” What does it mean to call a belief “reckless”?

According to one view, an actor who recklessly believes that *p* is one who believes that *p* but at the same time suspects that not-*p*. See, e.g., Douglas N. Husak & Craig A. Callender, *Wilful Ignorance, Knowledge, and the “Equal Culpability” Thesis: A Study of the Deeper Significance of the Principle of Legality*, 1994 *Wis. L. Rev.* 29, 41–42. Despite his belief that *p*, the actor's suspicion that not-*p* might provide the basis for requiring him to act to gather additional evidence, or simply to wait to acquire additional evidence, that would confirm his suspicion. According to another view, an actor who recklessly believes that *p* is one who believes that *p* while at the same time believing that he should believe that not-*p*. See, e.g., Larry Alexander, *Lesser Evils: A Closer Look at the Paradigmatic Justification*, 24 *Law & Phil.* 611, 624–25 (2005). This approach treats an actor who recklessly believes that *p* as someone suffering from epistemic akrasia. See *supra* note 29. For present purposes, I will continue to speak in the more familiar common-law terminology of reasonable and unreasonable belief.

39. According to one writer, self-defense is always an excuse. See Claire O. Finkelstein, *Self-Defense As a Rational Excuse*, 57 *U. Pitt. L. Rev.* 621, 643–44 (1996). According to

of considerable controversy. Some argue that such an actor, like the actor who reasonably and correctly believes that *p*, is justified in using deadly force.⁴⁰ Self-defense is therefore always a justification: An actor who kills because he reasonably believes that he is about to be killed has done nothing that the law does not permit him to do, whether his belief turns out to be correct or not.

Others argue that such an actor's use of deadly force is excused, but not justified: The law does not permit him to kill if as a matter of fact he was not about to be killed, despite his reasonable belief that he was about to be killed. Nonetheless, it withholds condemnation because his reasonable mistake makes it inappropriate to blame him for what he has done, even though it remains true that he should not have done it.⁴¹ Thus, self-defense is sometimes a justification; sometimes an excuse. It all depends on whether the actor's reasonable belief turns out to be true (justification) or not (excuse).

Likewise, nearly everyone agrees that an actor who unreasonably believes that *p* is neither justified nor (fully) excused.⁴² In some jurisdictions,

another, the entire excuse-justification debate is misguided inasmuch as the "restrictive schema of 'justification' and 'excuse' forces theorists to choose between just two alternative classifications, neither of which is satisfactory." R.A. Duff, *Rethinking Justifications*, 39 *Tulsa L. Rev.* 829, 838 (2004). Duff does not "discuss cases in which the actor's beliefs are unreasonable. . . ." *Id.* at 838 n.25.

40. See, e.g., Arthur Ripstein, *Equality, Responsibility, and the Law* 199 (1999); Victor Tadros, *Criminal Responsibility* 291 (2005); Marcia Baron, *Justifications and Excuses*, 2 *Ohio St. J. Crim. L.* 387, 387 (2005); Mitchell N. Berman, *Justification and Excuse, Law and Morality*, 53 *Duke L.J.* 1, 56 (2003); Russell L. Christopher, *Mistake of Fact in the Objective Theory of Justification: Do Two Rights Make Two Wrongs Make Two Rights . . . ?*, 85 *J. Crim. L. & Criminology* 295, 331 (1995); Joshua Dressler, *New Thoughts About the Concept of Justification in the Criminal Law: A Critique of Fletcher's Thinking and Rethinking*, 32 *UCLA L. Rev.* 61, 93 (1985); Kent Greenawalt, *The Perplexing Borders of Justification and Excuse*, 84 *Colum. L. Rev.* 1897, 1903 (1984); Kenneth W. Simons, *Self-Defense: Reasonable Beliefs or Reasonable Self-Control?*, 11 *New Crim. L. Rev.* 51, 65 (2008); Hamish Stewart, *The Role of Reasonableness in Self-Defense*, 16 *Can. J.L. & Jurisprudence* 317, 336 (2003).

41. See, e.g., Fletcher, *supra* note 29, § 10.1.2, at 766; Larry Alexander, *A Unified Excuse of Preemptive Self-Protection*, 74 *Notre Dame L. Rev.* 1475, 1483–84 (1999); John Gardner, *Justifications and Reasons, in Harm and Culpability* 103, 105 (A.P. Simester & A.T.H. Smith eds., 1996); Heidi M. Hurd, *Justification and Excuse, Wrongdoing and Culpability*, 74 *Notre Dame L. Rev.* 1551, 1564 (1999); Paul Robinson, *Competing Theories of Justification: Deeds v. Reasons, in Harm and Culpability, supra*, at 45, 47; Paul H. Robinson, *Criminal Law Defenses: A Systematic Analysis*, 82 *Colum. L. Rev.* 199, 239–40 (1982).

42. For claims contrary to this consensus, see Glanville Williams, *Criminal Law: The General Part* § 73, at 209 (2d ed. 1961) ("If a man inflicts injury on another in the unreasonable

including New York, such an actor would have no claim of self-defense to any charge. If charged with murder, he would be convicted of murder. In other jurisdictions, such an actor would have a defense to some charges, but not to others. If charged with murder, he would have a defense. But if charged with manslaughter, he would have none. Because his belief that *p* was unreasonable, the defense to which he is entitled is “imperfect.”⁴³ It constitutes a partial defense only. It does not result in an acquittal, but instead mitigates to manslaughter what would otherwise be murder.

The account developed here is similar to the doctrine of imperfect self-defense, but also different. According to the imperfect self-defense doctrine, or at least one prominent version of it, an actor who believes that *p*, but who unreasonably so believes, will be acquitted of murder, but convicted of voluntary manslaughter. Consequently, the doctrine is usually portrayed as *mitigating* murder to manslaughter.⁴⁴ The defendant is (really) guilty of murder, but the law reduces the crime from a greater one to a lesser one. Moreover, the mitigation the defense affords remains the same (murder to manslaughter), no matter why the actor’s belief that *p* was unreasonable.

In contrast, on the account developed here, an actor who believes that *p* should, all else being equal, be acquitted on grounds of self-defense. He should be excused. He has in fact acted impermissibly, but he is not blameworthy. Looking at the world through his eyes, he has at the moment he kills done nothing that the law itself would not have permitted him to do. Nonetheless, he loses that defense if he believed that *p* only because he culpably violated some other obligation with respect to which

belief that he has to do so in self-defence, the injury must be borne philosophically as an accident. . . . If this is so, why should the position be any different where the act of supposed self-defense results in death?”; Singer, *supra* note 22, at 461 (concluding that the “subjective test is preferable to the objective rule [i.e., the reasonable-belief rule] courts embraced in the nineteenth century”).

43. See, e.g., Dressler, *supra* note 19, § 18.03, at 249; LaFave, *supra* note 19, § 5.7(i), at 500–01. At least one study finds that most people, given a choice, would choose to make an unreasonably mistaken actor’s liability proportionate to the culpability associated with his mistake. See Paul H. Robinson & John M. Darley, *Testing Competing Theories of Justification*, 76 N.C. L. Rev. 1095, 1128 (1998).

44. See, e.g., Dressler, *supra* note 19, § 18.03, at 249–50 (noting that imperfect self-defense mitigates culpability); Robinson, *supra* note 38, § 8.5, at 460 (noting that MPC version of imperfect self-defense mitigates degree of liability).

the state can legitimately demand compliance. Thus, whereas the doctrine of imperfect self-defense treats an honest (but unreasonable) belief as mitigating, the account developed here treats an unreasonable (but honest) belief as *aggravating*. Moreover, whereas the doctrine of imperfect self-defense provides a one-size-fits-all mitigation, the account developed here would ideally tailor the extent of the forfeiture to match the seriousness of the prior breach. Relatively more serious breaches should result in more extensive forfeitures (and more extensive liability), and vice versa.

Let me begin the development of this account with a more general discussion of the role forfeiture rules play in the criminal law.

A. Forfeiture Rules

Criminal law defenses like self-defense often come with strings attached. These strings take the form of forfeiture rules. Under these rules an actor forfeits a defense to which he would otherwise have been entitled if he culpably chooses to act or not act at time t_1 , which act or omission is the but-for and proximate cause of his being subject at time t_2 to the type of threat associated with the relevant defense.⁴⁵ The general idea behind such rules is that an actor who culpably chooses to create or encounter a threat, whether that threat comes from man or nature, should not be allowed to point to that threat if, should the threat come to pass, he is forced to commit a crime in order to avoid it. He forfeits (in full or in part) any excuse

45. See, e.g., 2 Robinson, *supra* note 19, § 123(a), at 30 (“[A]ll jurisdictions with law on this point take into account the actor’s culpability in causing or contributing to the justifying circumstances, and limit the availability of the justification defenses.”); *id.* § 162(a), at 247 (noting that although “the problem of an actor causing his disability or condition most frequently arises in cases involving intoxication . . . [t]here is no reason . . . why such a circumstance should not be taken into account for all excuses”).

Some forfeiture rules are based, not on the actor’s culpable choice to encounter a threat, but instead on his negligent encountering of it. For example, under the Model Penal Code an actor who “was negligent in placing himself in . . . a situation [in which it was probable that he would be subjected to duress]” forfeits any claim of duress “whenever negligence suffices to establish culpability for the offense charged.” Model Penal Code § 2.09(2) (Proposed Official Draft 1962). Because I believe that negligence is a controversial basis upon which to premise a forfeiture rule, I set aside such negligence-based rules for present purposes. I believe that negligence is a controversial basis upon which to premise a forfeiture rule because I believe that negligence is (most often) an illegitimate basis upon which

or justification to which he would otherwise have been entitled based on his choice to create or encounter the threat in the first place. He violates a duty against culpably choosing to create or encounter threats, and the price he pays for that choice is to lose (in full or in part) a defense to which he would otherwise have been entitled.

The so-called “aggressor rule” associated with self-defense is a good example. According to the Model Penal Code’s formulation, an actor who “provoke[s] the use of force against himself in the same encounter”⁴⁶ forfeits any claim of self-defense to which he might otherwise have been entitled, provided that he provoked the use of such force against himself “with the purpose of causing death or serious bodily injury.”⁴⁷ In other words, if you throw a punch in order to provoke someone to try to kill you so that you can kill him first, and if he does try to kill you, you cannot claim self-defense if you kill him before he kills you. You are the initial aggressor, and as such you forfeit your right to self-defense, which you can regain only if you renounce your initial aggression.

Similar forfeiture rules often accompany the defenses of necessity and duress.⁴⁸ For example, the Model Penal Code provides that an actor who commits a crime under duress, and who would therefore otherwise have a valid defense to the crime charged, forfeits that defense if he “recklessly placed himself in a situation in which it was probable that he would be

to premise retributive punishment, and forfeiture rules end up imposing retributive punishment for the act or omission forming the basis for the forfeiture.

46. Model Penal Code § 3.04(2)(b)(i) (Proposed Official Draft 1962).

47. *Id.* The real Goetz would not have lost his claim of self-defense under this provision. Even if he did in fact do something to “provoke the use of force against himself,” nothing in the available facts suggests that he did so with the “purpose of causing death or serious bodily injury.” I assume, for example, that even if Goetz’s choice to sit close to the four youths provoked (i.e., was a but-for cause of) their use of force against him, Goetz did not make that choice in order to cause death or serious bodily injury to the victims. The result might be different under broader (and therefore more controversial) formulations of the aggressor rule.

48. Forfeiture rules are for some reason seldom attached to insanity-defense provisions. See, e.g., Paul H. Robinson, *Causing the Conditions of One’s Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine*, 71 Va. L. Rev. 1, 24 & n.85 (1985) (identifying only two states whose penal codes deny an actor an insanity defense when the actor culpably chooses to cause his own insanity). Although an actor may bear no responsibility for being mentally diseased or defective, he may nonetheless bear some responsibility under some circumstances if he permits his mental disease or defect to cause the cognitive

subjected to duress.”⁴⁹ Thus, an actor who joins a crime-committing gang and then finds himself forced to commit a crime cannot claim duress as a defense inasmuch as he culpably chose to place himself in the situation giving rise to the duress. Much the same goes for necessity.⁵⁰ Under rules of this type, an actor forfeits a defense to which he would otherwise have been entitled if and because he culpably impairs his *situation*. He chooses to get himself into trouble.

The voluntary-intoxication rule is another type of forfeiture rule.⁵¹ Getting drunk, like starting a fight or joining a gang, can cost an actor a defense to which he would otherwise have been entitled. For example, if an actor unwittingly creates an unjustified risk of causing someone’s death,

or volitional incapacity associated with traditional tests of insanity; for example, choosing not to take medicine he knows that he needs to take in order to control the effects of his disorder. See, e.g., Michael D. Slodov, Note, Criminal Responsibility and the Noncompliant Psychiatric Offender: Risking Madness, 40 Case W. Res. L. Rev. 271, 274 (1990) (arguing that “in some circumstances, imposing responsibility on the noncompliant mentally ill offender is consistent with the aims of criminal law and with accepted principles of criminal responsibility”).

49. Model Penal Code § 2.09(2) (Proposed Official Draft 1962). The defense is forfeited completely if the actor recklessly placed himself in such a situation (and presumably if he does so purposely or knowingly as well), but only partially if the actor negligently places himself in such a situation. See *id.*

50. See, e.g., Model Penal Code § 3.02(2) (“When the actor was reckless or negligent in bringing about the situation requiring a choice of harms or evils. . . , the justification afforded by this Section is unavailable in a prosecution for any offense for which recklessness or negligence, as the case may be, suffices to establish culpability.”). An actor who purposely or knowingly brought about the situation requiring such a choice would presumably lose the defense altogether.

51. This rule is sometimes characterized as an evidentiary rule and sometimes as a substantive rule. Take the case of a drunken defendant who unwittingly kills someone and is charged with reckless homicide, which requires awareness of the lethal risk his conduct is creating. Characterizing the voluntary-intoxication rule as an evidentiary rule would mean that the state is required to prove that the defendant realized that he was creating a lethal risk, but that the defendant is prevented from introducing intoxication evidence designed to show that he lacked the requisite awareness. Characterizing the rule as a substantive rule would mean that the state is not required to prove the requisite awareness. Instead, it would mean that the state believes that getting drunk and unwittingly causing death is just as serious as the crime of reckless homicide (consciously imposing an unjustified lethal risk with death resulting). Compare *Montana v. Egelhoff*, 518 U.S. 37, 41 (1996) (plurality opinion) (characterizing the Montana voluntary-intoxication rule at issue in the case as a rule of

and someone gets killed, he is not guilty of reckless homicide, because reckless homicide requires the state to prove that he *was* aware of the lethal risk he was creating. But if the reason the actor failed to realize that he was creating a lethal risk is because he got himself drunk, then he is out of luck. Though he was not in fact reckless, the law treats him as if he was.⁵² He chose to drink, and unlucky for him, he happened to kill someone while intoxicated, even though he never realized that he was exposing anyone to a risk of death. Under the voluntary-intoxication rule an actor forfeits a defense to which he would otherwise have been entitled if and because he culpably impairs his *mind* through intoxicants. He chooses to become unaware.⁵³

Forfeiture rules are objectionable because they convict and punish an actor for a crime he did not commit or to which he would otherwise have

evidence excluding evidence of the effects of voluntary intoxication), with *id.* at 57 (Ginsburg, J., concurring) (characterizing the rule as a substantive rule redefining the mental-state element of the offense charged). See also Peter Westen, *Egelhoff Again*, 36 *Am. Crim. L. Rev.* 1203, 1215–27 (1999) (describing these two approaches).

Some jurisdictions treat mental-illness evidence in a manner analogous to voluntary-intoxication evidence. In these jurisdictions mental-illness evidence can be introduced to show that the defendant was insane, but it cannot be introduced to show that he lacked a mental state associated with the crime charged. See Dressler, *supra* note 19, § 26.02[B][4], at 397–98. Although this mental-illness rule may be defended on a variety of evidentiary grounds, see, e.g., *Clark v. Arizona*, 126 S. Ct. 2709, 2734–36 (2006), it would be harder to defend on substantive grounds. Treating a voluntarily intoxicated actor as if he possessed a mental state he did not in fact possess is one thing: A voluntarily intoxicated actor is at least responsible for becoming intoxicated. Treating a mentally ill actor as if he possessed a mental state he did not in fact possess is another: A mentally ill actor is ordinarily responsible neither for becoming mentally ill nor for the behavioral manifestations of his illness.

52. See, e.g., Model Penal Code § 2.08 (Proposed Official Draft 1962). For three slightly different interpretations of § 2.08, see Westen, *supra* note 51, at 1220 n.72. A voluntarily intoxicated actor would still have a failure-of-proof defense to a charge of purposeful or knowing homicide (denominated murder) under the MPC. Less clear is whether the actor would continue to have such a defense to a charge of reckless homicide under circumstances manifesting extreme indifference to the value of human life (also denominated murder).

53. Opponents of the voluntary-intoxication rule have proposed a separate crime of “being drunk and dangerous” for which “conviction . . . should usually result in purely remedial treatment . . . [and] could even result in punishment if the accused, knowing from previous experience that he is dangerous when in liquor, continues to take it.” Williams, *supra* note 42, § 183, at 573–74.

had a valid defense. The “crime” the actor actually committed was that associated with his prior culpable choice: joining a gang, getting drunk, and so forth. The punishment the actor deserves is whatever punishment (if any) is deserved for making that choice. The actor who joins a criminal gang, hoping or believing he will or might later be coerced into committing a robbery, should be punished for choosing to join the gang with those attendant mental states, not for the robbery he committed under duress. The actor who gets drunk and unwittingly kills someone should be punished for getting drunk and unwittingly causing death, not for reckless homicide. Actors should be punished for the crimes they commit, and the punishment they receive should fit the crime. But forfeiture rules result in actors being punished for crimes they did not commit, and as such, forfeiture rules result in disproportionate punishment, though *how* disproportionate will of course depend on the facts.

Nonetheless, my goal here is not to criticize forfeiture rules, nor urge that they be banished from the criminal law.⁵⁴ Instead, I want to urge that the reasonable-belief rule of self-defense is fairly portrayed as a forfeiture rule, and that a liberal state cannot legitimately apply this rule in cases like that of Goetz*.

B. The Reasonable-Belief Rule As a Forfeiture Rule

Although the reasonable-belief rule is not usually portrayed as a forfeiture rule, it seems to me that it can fairly be so portrayed. The reasonable-belief rule is like the voluntary-intoxication rule inasmuch as both involve an impairment of the actor’s mind. The intoxicated actor’s impairment takes the form of ignorance: He fails to form a belief that he should have formed, and but for his intoxication would have formed. He fails to see something that he should have seen. The unreasonably believing actor’s impairment takes the form of a mistake: He forms a belief that he should not have formed, and but for some prior breach of duty, would not have formed. He sees something that he should not have seen.

A voluntarily intoxicated actor forfeits a defense because he chose to get drunk. An unreasonably believing actor forfeits a defense because he unreasonably believes, and in the case of Goetz*, because he unreasonably believes

54. See Robinson, *supra* note 48, at 28–29.

that *p*.⁵⁵ But what does it mean to say that Goetz* unreasonably believes that *p*? What is the something that provides the basis upon which he forfeits his claim of self-defense? If Goetz*'s belief that *p* was unreasonable, such that he forfeits his claim to self-defense, *why* was it unreasonable? In order to answer that question, we need a better sense of what was going on in Goetz*'s head at the moment he pulled the trigger. What was he thinking?

Here is one account. We are assuming that at some point during the fatal encounter, Goetz* formed the belief that his life was in imminent danger, which belief may or may not have had any thought preceding it. That belief was a belief about the world: a belief about what is the case. That belief in turn caused, or may have caused, Goetz* to think about what he ought to do. He then formed another pair of beliefs: that he ought to save himself from his assailant's imminent attack, and that in order to do so he ought to kill his assailant. These beliefs are practical judgments: beliefs about what one ought to do. At that point, consistent with his judgment as to what he ought to do, Goetz* chose to kill his assailant, thereby causing himself to form the intent to kill. He then chose to execute that intention, resulting in the formation of a volition, which in turn caused his finger to move and the trigger to be pulled. The rest was up to the laws of nature.

The belief that sets this sequence in motion is the belief that *p*. Yet that belief, like any other belief an actor forms at any given moment, depends on the evidence available to him at that moment, as well as on his cognitive capacities at that moment. For present purposes, I assume that nothing was wrong with Goetz*'s cognitive capacities. As such, I assume that his formation of the belief that *p* was not due to anything that could fairly be characterized as a defect in cognitive capacity or mental disorder, such as "racial paranoia."⁵⁶ Instead, I assume that the problem was with his evidence. The problem, one might say, was not with Goetz*'s cognitive hardware, but with

55. One might argue that an actor subject to the reasonable-belief rule would not otherwise have a valid defense, whereas an actor subject to the voluntary-intoxication rule would. An actor who kills because he unreasonably believes that he is about to be killed does not have a valid self-defense claim, so the argument goes, because a valid self-defense claim requires his belief to be reasonable. But this argument would seem to beg the question. It simply presupposes that the reasonable-belief rule is somehow intrinsic to the defense itself when the rule can also be portrayed as a forfeiture rule extrinsic to it.

56. I assume that Goetz* did not believe that *p* just because he believed that his putative assailants were black; that *would* be paranoid. Cf. American Psychiatric Association,

his software. Finally, I will assume that the evidence available to Goetz* at the moment he formed the belief that *p* consisted of all the other beliefs he possessed at that moment. Call these beliefs his background beliefs: beliefs he possessed at the moment he formed the belief that *p* and but for which he would not have formed the belief that *p*.

Goetz*'s background beliefs no doubt included beliefs related to the victim's movements or gestures, his request or demand for five dollars, the tone of his voice, the look in his eye, the tight confines of the subway car, and so forth. None of these beliefs is thought to be particularly objectionable. They are legitimate grounds upon which anyone might form the belief that *p*. We can also assume that Goetz*'s background beliefs included two other beliefs: that males are more prone to violence than females; and that the young are more prone to violence than the old. These beliefs are of course generalizations.⁵⁷ Even so, I doubt that many people would consider them illegitimate bases upon which one might form the belief that *p*.

If all these background beliefs were jointly sufficient to have caused Goetz* to form the belief that *p*, then the case would be far less interesting than it would be if they were insufficient. What makes the case interesting is the assumption that Goetz* believed that *p* because and only because his network of background beliefs included another generalization: blacks are more prone to violence than nonblacks, or some proposition along those lines. Call this belief the belief that *q*. I will assume that this belief was necessary to Goetz*'s formation of the belief that *p*, and that it, together with his other background beliefs, were sufficient to cause him to form the belief that *p*. Goetz*'s possession of the belief that *q* is usually what leads to his characterization as a racist, and inasmuch as his racism consisted in his possession of that belief, Goetz*'s racism was cognitive.⁵⁸ It was, so to speak, in his head.

Diagnostic and Statistical Manual of Mental Disorders 690 (4th ed., textual rev. 2000) ("Individuals with [paranoid personality] disorder assume that other people will exploit, harm, or deceive them, even if no evidence exists to support this expectation. . . .").

57. In fact all of his relevant background beliefs are generalizations, i.e., people who make gestures like the gestures the victim made are more prone to violence; people who make requests or demands for money are more prone to violence; and so forth.

58. See, e.g., Kwane Anthony Appiah, *Racisms*, in *Anatomy of Racism* 3, 5 (David Theo Goldberg ed., 1990) ("[E]xtrinsic racisms make moral distinctions between members of different races because they believe that racial essence entails certain morally relevant qualities.").

Commentary on the real Goetz case tends to suggest that Goetz*'s racism was indeed cognitive.⁵⁹ Goetz* was a racist because he *believed* that *q*. But there is another possibility. Goetz*'s racism may not have been in his head, but in his heart. In other words, his racism was conative,⁶⁰ not cognitive.⁶¹ Conative racism can be a matter of hostility, ill will animus, malice, and so forth, in which case the actor wants members of the stigmatized group to suffer some disadvantage or bear some burden,⁶² just because they are members of the stigmatized group; or it can be a matter of indifference, in which case the actor cares not at all, or less than he should, for the well-being or fate of the group's members.⁶³ Call this desire, or relative lack

59. See, e.g., Armour, *supra* note 36, at 782 (racism consists in the actor's belief that "blacks are more prone than whites to be criminals"); Kelman, *supra* note 36, at 812 (racism consists in the actor's beliefs "about the criminal predilections of black teenagers").

60. See, e.g., J.L.A. Garcia, *The Heart of Racism*, 27 *J. Soc. Phil.* 5, 6 (1996) ("Racism . . . is something that essentially involves not our beliefs and their rationality or irrationality, but our wants, intentions, likes and dislikes. . . .") [hereinafter Garcia, *Heart of Racism*]. Garcia has defended this account against competitors in subsequent work. See, e.g., J.L.A. Garcia, *Current Conceptions of Racism: A Critical Examination of Some Recent Social Philosophy*, 28 *J. Soc. Phil.* 5 (1997); J.L.A. Garcia, *Philosophical Analysis and the Moral Concept of Racism*, 25 *Phil. & Soc. Criticism* 1 (1999); J.L.A. Garcia, *Racism and Racial Discourse*, 32 *Phil. F.* 125 (2001). For criticism of Garcia's conative conception of racism, see, e.g., Charles W. Mills, "Heart" Attack: A Critique of Jorge Garcia's Volitional Conception of Racism, 7 *J. Ethics* 29, 44 (2003) ("An account of racism which just focuses on feelings without an examination of their accompanying beliefs is not going to work because we need to know what beliefs *ground* the feelings in order to adjudicate whether they are racist or not."); Tommie Shelby, *Is Racism in the "Heart"?*, 33 *J. Soc. Phil.* 411, 414 (2002) (arguing that racist "beliefs are essential to and even sufficient for racism").

61. For another take on the distinction between cognitive and conative racism, see Lawrence Blum, "I'm Not a Racist, But . . .": The Moral Quandary of Race 8 (2002) (distinguishing between "inferiorization" (cognitive) and "antipathy" (conative) racism).

62. See, e.g., Garcia, *Heart of Racism*, *supra* note 60, at 6 ("In its central and most vicious form, [racism] is a hatred, ill-will, directed against a person or persons on account of their assigned race.").

63. See, e.g., *id.* ("In a derivative form, one is a racist when one either does not care at all or does not care enough (i.e., as much as morality requires) or does not care in the right ways about people assigned to a certain racial group, where this disregard is based on racial classification.").

A number of criminal-law scholars have argued that indifference, whether race-based or otherwise, does and should play an important role in criminal-law theory and doctrine. For example, they have argued that an actor who creates a risk of causing a prohibited harm, but who does so unwittingly, can still fairly be subject to retributive punishment if

of desire, the desire that q .⁶⁴ Cognitive racism and conative racism may go hand in hand. Racial animus may cause an actor to hold a racist belief (which explains why such beliefs tend to be impervious to countervailing evidence); and racial beliefs may cause an actor to harbor racial animus. But they can also travel separately. An actor might believe that blacks are more violent than nonblacks without harboring any malice toward them; or he might harbor malice toward them without possessing

his lack of awareness was due to indifference to the well-being of others. See, e.g., R.A. Duff, *Intention, Agency, and Criminal Liability* 157 (1990) (Culpable negligence is “essentially a matter . . . of a kind of ‘practical indifference.’”); Mayo Moran, *Rethinking the Reasonable Person: An Egalitarian Reconstruction of the Objective Standard* 258 (2003) (“[T]he indifference account places its focus on the attitude displayed by any particular action. . . .”); Samuel H. Pillsbury, *Judging Evil: Rethinking the Law of Murder and Manslaughter* 171 (1998) (“Where the accused did not perceive the risks involved at the time of his conduct, culpability rests on a judgment about why the person failed to perceive.”); Jeremy Horder, *Gross Negligence and Criminal Culpability*, 47 *U. Toronto L.J.* 495, 501 (1997) (“The subjective element in indifference lies . . . in an uncaring attitude toward the victim’s relevant protected interests.”); Samuel Pillsbury, *Crimes of Indifference*, 49 *Rutgers L. Rev.* 105, 151 (1996) (“The key to culpability for failure to perceive is why the person failed to perceive.”); Kenneth W. Simons, *Does Punishment for “Culpable Indifference” Simply Punish for “Bad Character”? Examining the Requisite Connection Between Mens Rea and Actus Reus*, 6 *Buff. Crim. L. Rev.* 219, 264 (2002) (One “possible culpable indifference standard . . . asks what the actor would have done if he had had a different belief about the relevant risks.”); Kenneth W. Simons, *Rethinking Mental States*, 72 *B.U. L. Rev.* 463, 487 (1992) (“[R]eckless indifference . . . [means] caring much less about the result than the actor should.”); Kenneth W. Simons, *Culpability and Retributive Theory: The Problem of Criminal Negligence*, 5 *J. Contemp. Legal Issues* 365, 388 (1994) (“Culpable indifference . . . is a desire-state reflecting the actor’s grossly insufficient concern for the interests of others.”); Victor Tadros, *Recklessness and the Duty to Take Care*, in *Criminal Law Theory* 227, 229 (Stephen Shute & A.P. Simester eds., 2001) (arguing that criminal liability for negligence is not warranted unless the “defendant’s action is a manifestation of one of a narrow range of vices: primarily, vices that show that the defendant has insufficient regard for the interests of others”). For criticism of this line of thought, see, e.g., Larry Alexander, *Insufficient Concern: A Unified Conception of Criminal Culpability*, 88 *Cal. L. Rev.* 931, 938 (2000); Stephen P. Garvey, *What’s Wrong With Involuntary Manslaughter?*, 85 *Tex. L. Rev.* 333, 357–63 (2006).

64. Describing an actor who is indifferent to the well-being of blacks as possessing the desire that q is of course not quite right. It would be more precise to say that he lacks sufficient desire to treat blacks with the equal concern and respect to which everyone is entitled.

any belief or set of beliefs that might rationalize or make sense of such a sentiment.⁶⁵

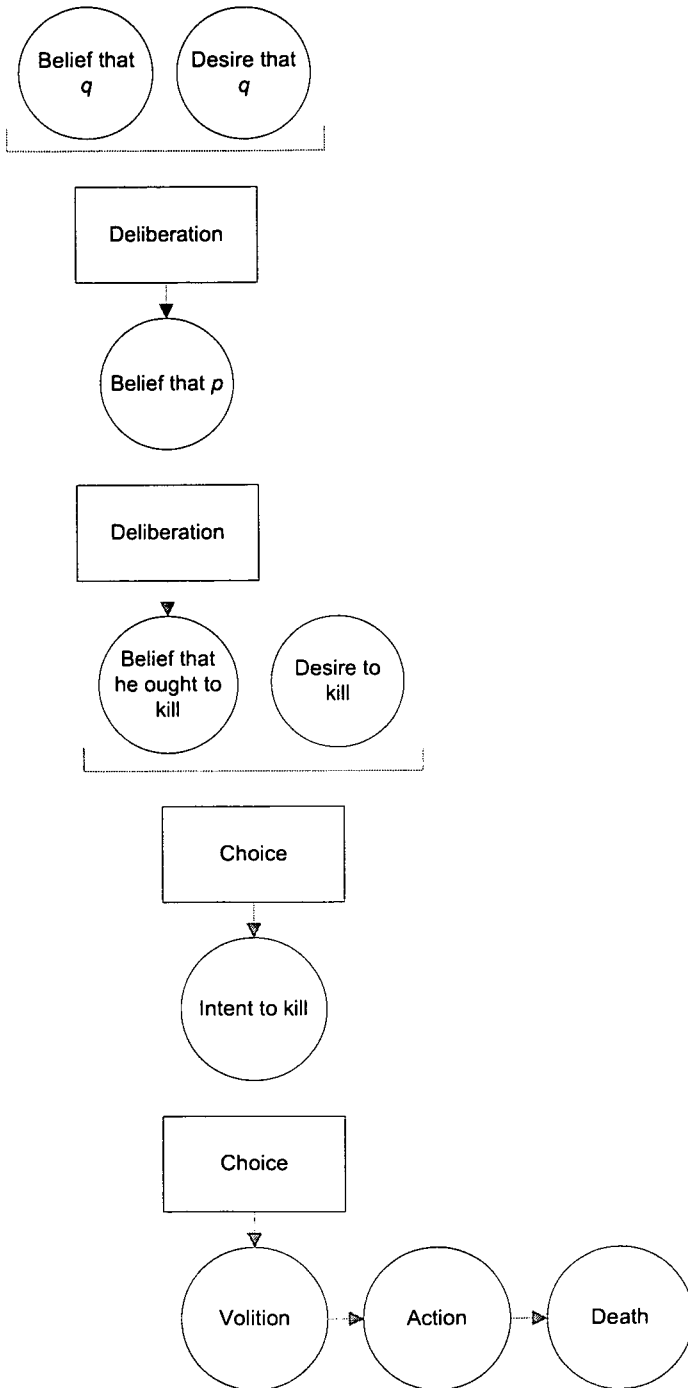
The important point for now is that the beliefs an actor forms at any moment in time depend not only on the background beliefs he possesses at that moment, but also on the background desires he possesses. When some desire other than the desire to discover the truth and avoid error influences what an actor believes, we might describe the actor as self-deceived, or say that his belief is the product of wishful thinking.⁶⁶ Indeed, we use such labels in order to capture the causal power desire can have on belief formation. If Goetz* was a conative racist, then any background desire to discover the truth and avoid error was not the only desire influencing what he believed. His animus or indifference toward blacks may also have caused him to believe what he did, and for present purposes I will assume that they did. Thus, but for his possession of the desire that *q*, or the belief that *q*, Goetz* would not have formed the belief that *p*. He formed the belief that *p* because and only because he was a racist, either cognitive or conative.

We can depict the foregoing snapshot of Goetz's folk or commonsense psychology as follows.⁶⁷

65. See, e.g., Blum, *supra* note 61, at 10–11 (“Inferiorizing and antipathy racism are distinct. Some inferiorizing racists do not hate the target of their belief. . . . Conversely, not every race hater regards the target of her hatred as inferior.”).

66. See, e.g., Robert Audi, *Self-Deception, Rationalization and the Ethics of Belief: An Essay in Moral Psychology*, in Robert Audi, *Moral Knowledge and Ethical Character* 131, 132 (1997) (offering an analysis of self-deception according to which among other things an actor is self-deceived if, having formed a true belief, desire pushes that belief into unconsciousness, such that the actor sincerely avows that which is false); Alfred R. Mele, *Self-Deception Unmasked* 50–51 (2001) (offering an analysis of self-deception according to which an actor is self-deceived if among other things his desires cause him to form a false belief); Béla Szabados, *Wishful Thinking and Self-Deception*, 33 *Analysis* 201, 204 (1973) (claiming that the wishful thinker and the self-deceiver share in common the fact that “[b]oth hold the belief they do hold largely because they want to believe” as they do).

67. For a recent defense of the criminal law's dependence on such psychology against some of the challenges from various forms of eliminativism, see Katrina L. Sifferd, *In Defense of the Use of Commonsense Psychology in the Criminal Law*, 25 *Law & Phil.* 571 (2006).



With this account in hand, we can formulate three theories according to which Goetz*'s belief that *p* was unreasonable. According to the first, it was unreasonable for Goetz* to believe that *p* because his racist belief or desire that *q* caused him to believe that *p*, and he should not have possessed that racist belief or desire. According to the second, it was unreasonable for Goetz* to believe that *p* because he chose to believe that *p*; he could have chosen otherwise, and he should have chosen otherwise inasmuch as his choice to believe that *p* was based on the racist belief or desire that *q*. According to the third, it was unreasonable for Goetz* to believe that *p*, not because he was a racist and his racism caused him to believe that *p* (the first theory), nor because he chose to believe that *p* and his choice was based on racism (the second theory). Instead, his belief that *p* was unreasonable because his racism caused him to form the belief that *p*, and he chose to become or remain a racist when he could and should have chosen otherwise.

The first theory (discussed in part II) links the unreasonableness of Goetz*'s belief that *p* directly to his racist character. He forfeits his claim to self-defense because he was a racist. The second and third theories (discussed in parts III and IV) link the unreasonableness of Goetz*'s belief that *p* to a choice he made or failed to make, either to form the belief that *p*, or to become or remain a racist. He forfeits his claim to self-defense because he made a choice he should not have made.

II. THE CHARACTER THEORY

The first theory of unreasonableness claims that Goetz*'s belief that *p* was unreasonable because his racism caused him to form that belief, and he should not have been a racist. On this theory, Goetz* loses his claim of self-defense because of the content of his character, because he possessed racist beliefs or desires. He loses the defense because of who he is: a racist.

A liberal state can embrace the character theory of unreasonableness if two conditions are satisfied. First, a liberal citizen ought not to possess the racist beliefs or desires we are assuming that Goetz* possessed. Second, a liberal state must be at liberty to rely upon the standing beliefs or desires an actor possesses as a basis upon which to deny him a defense to which he would otherwise have been entitled. In other words, it must be the case that a liberal state can rely upon the content of an actor's character as the basis for a forfeiture rule. I argue that the first condition is satisfied, but not the second. Consequently, a liberal state cannot embrace the character theory of unreasonableness.

A. Is Racism Wrong?

Racism can be rooted in an actor's beliefs, or in his desires, or both. Racism rooted in an actor's desires (racism in the heart) is straightforwardly inconsistent with citizenship in a liberal state. Racism rooted in an actor's beliefs (racism in the head) bears a more complicated relationship to the ideal of liberal citizenship.

1. Racism in the Heart

An actor whose racism resides in his heart is one who harbors animus, hostility, ill will, and so forth toward blacks, or who is at least indifferent to them. Such animus or indifference can enter into an actor's psychology in two very different ways. First, racial animus can directly influence what an actor does. Its expression in action may be the point or purpose of the action, or at least part of its point or purpose. Second, racial animus or indifference can influence what an actor believes, and thereby indirectly influence what he does.

When animus manifests itself directly in action, and when that action already constitutes a criminal offense, what would otherwise be a run-of-the-mill crime turns into a hate crime.⁶⁸ Although the subject is one of considerable controversy, a hate crime is probably best analyzed as an ordinary crime committed with a specific goal or purpose in mind.⁶⁹ A criminal act

68. An actor can of course choose to express his racial hatred or animus in acts that are not already crimes, including speech acts. Expressions of racial hatred—whether through speech acts or other forms of action—are fair targets of criminalization insofar as an actor has control over whether or not to engage in them. Indeed, insofar as an actor has control over the formation of an intent to humiliate or degrade another person, the formation of such an intent might itself be a fair target of criminalization, all else being equal. Nonetheless, a range of countervailing considerations, including those values associated with the First Amendment, counsel against the criminalization of such speech acts or acts of intent formation. Compare *R.A.V. v. City of St. Paul*, 505 U.S. 377, 392 (1992) (holding that an ordinance criminalizing “fighting words that contain . . . messages of ‘bias-motivated’ hatred” violates the First-Amendment rule against content-based discrimination), with *Wisconsin v. Mitchell*, 508 U.S. 476, 490 (1993) (holding that a penalty enhancement for a defendant who selected his victim because of the victim's race does not violate the First Amendment).

69. See, e.g., Kent Greenawalt, Reflections on Justifications for Defining Crimes by the Category of Victim, 1992 *Ann. Surv. Am. L.* 617, 620–25; Jeffrie G. Murphy, Bias Crimes: What Do Haters Deserve?, 11 *Crim. Just. Ethics*, Summer/Fall 1992, at 20, 21; Paul H. Robinson, Hate Crime: Crime of Motive, Character, or Group Terror?, 1992/1993 *Ann. Surv. Am. L.* 605, 606–09. But see Heidi M. Hurd & Michael S. Moore, Punishing Hatred and

turns into a hate crime when committed with the specific intent or purpose to humiliate, degrade, or otherwise insult the victim because he is a member of a protected class. The victim's humiliation is the end toward which the actor acts, or at least one of his ends. It supplies the motive for the action, and the actor wholeheartedly embraces that motive. On this view of what makes a hate crime a crime of hate, Goetz*'s actions cannot be so portrayed, inasmuch as his motivation was self-preservation, not humiliation. He killed his putative assailant because he believed that his putative assailant was about to kill him, not in order to humiliate or degrade.

Besides being expressed directly in action, racial animus or indifference can also influence action indirectly, exercising its force in the first instance on the beliefs an actor forms. An actor who harbors racial animus or indifference is apt to form beliefs his non-racist counterpart would not form, and conversely, he is apt not to form beliefs his non-racist counterpart would form. We might say that such an actor is one whose desire that *q* "acts on" him, whereas an actor who commits a hate crime is one who "acts on" his desire that *q*. Thus, Goetz*'s desire that *q* may have caused him to form the belief that *p*, which ultimately caused him to form the intent to kill. He did not endorse and express that desire in action, but that desire nonetheless caused him to form a belief he would not otherwise have formed.

Was it wrong for Goetz* to possess the desire that *q*, even if he did not directly act on it? If we assume that most people possess the desire that *q*, at least to one degree or another, then one might argue that it was not wrong. In other words, one might say that the reasonable person *is* a conative racist, because the reasonable person is the typical or ordinary person, and regrettably, the typical person is a conative racist.⁷⁰ If so, then it was not wrong for Goetz* to harbor animus toward blacks, or at least be indifferent to them.

Perhaps, but the better view is that citizens of a liberal state ought not to possess the desire that *q*, whether or not most of them in fact do.⁷¹ The simple fact of the matter is that a liberal citizen does not harbor race-based

Prejudice, 56 Stan. L. Rev. 1081, 1123 (2004) (arguing that "hate/bias crimes concern themselves with new and novel sorts of mens rea" that cannot be understood as a form of specific intent).

70. See, e.g., Armour, *supra* note 36, at 787 ("The Reasonable Racist asserts that, even if his belief that blacks are 'prone to violence' stems from pure prejudice, he should be excused for considering the victim's race before using force because most similarly situated Americans would have done so as well.").

71. See, e.g., Lee, *supra* note 28, at 235 ("[N]ormative reasonableness is a conception of reasonableness that focuses on the beliefs and actions society ought to recognize as reasonable.

animus toward his fellow citizens, nor does he harbor race-based indifference. The conative constitution of liberal citizens has no room for such sentiments. As a citizen of a liberal state, Goetz* should not have possessed the desire that *q*. Moreover, had he not possessed that desire, he would not have formed the belief that *p*, and had he not formed the belief that *p*, he would not have pulled the trigger.

2. Racism in the Head

Goetz* should not have possessed the desire that *q*. But whether he should or should not have possessed the belief that *q*—blacks are more prone to violence than nonblacks—turns out to be more controversial. Racism in the heart is out-of-bounds for liberal citizens, but what about racism in the head? At the outset, we should reject once again the idea that Goetz* should have believed that *q*, or was at least permitted to believe that *q*, because most people believe that *q*, assuming that they do.⁷² What most people believe is neither here nor there. A widespread belief might nonetheless be one liberal citizens should not hold.

Instead, the argument to the effect that Goetz* ought to have believed that *q*, or was at least permitted to believe that *q*, boils down to the claim that *q* is true, and one ought to believe the truth, not to mention be permitted to believe it. Moreover, one can hardly be called a racist for believing the truth. One can imagine two responses to this argument.

The first response is that the proposition that *q* is *not* true; on the contrary, it is false. It purports to be a valid statistical generalization when in fact it is not.⁷³ Instead, it is a false or misleading generalization, or in other

A positivist (or empirical) conception of reasonableness, in contrast, focuses on what most individuals would actually feel, think, or do if they were in the defendant's situation.”)

72. According to a poll taken around the time of the real Goetz case, “[w]hen black New Yorkers were asked whether they would feel unsafe if they saw several loud, teenage white boys on their subway car, 39 percent said yes. Would they feel similarly unsafe if the youths were black? Yes, 51 percent said. The responses by whites was 55 percent and 71 percent, respectively.” Sam Roberts, *Exploring Laws and the Legacy of the Goetz Case*, N.Y. Times, Jan. 23, 1989, at B1.

73. Compare Armour, *supra* note 36, at 792 (“Even if we accept the . . . claim that his greater fear of blacks results wholly from his unbiased analysis of crime statistics, biases in the criminal justice system undermine the reliability of the statistics themselves.”), with Randall Kennedy, *Suspect Policy*, New Republic, Sept. 13 & 20, 1999, at 30, 32 (“Statistics abundantly confirm that African Americans—and particularly young black men—commit

words, a stereotype.⁷⁴ Or, even if the generalization itself is statistically valid, it tends to exercise undue influence on a person's thought processes, getting more weight than it deserves, such that it ends up operating as a *de facto* stereotype.⁷⁵ In either case, if a proposition is false, or if it otherwise corrupts an actor's belief-formation process so as to cause him to form other beliefs that are false, then the actor should not believe the offending proposition. Thus, Goetz* should not have believed that *q*.

The second response accepts or concedes that *q* is true, and it acknowledges that an actor cannot be faulted, all else being equal, for believing that which is true, but it nonetheless insists that all else is not equal. Even if the proposition that *q* is true, even if it *is* a valid statistical generalization, and not a stereotype, one might nonetheless argue that we should *not* always believe the truth, nor therefore should the truth always be permitted to influence other beliefs we form.⁷⁶ We might all be better off believing that *q* is false, even if *q* is true. Thus, Goetz* should not have believed that *q*.

Consider the debate over racial profiling. For example, suppose that it turns out to be true that black motorists driving along a certain stretch of highway are, all else being equal, more apt to be carrying contraband than are nonblack motorists. Call this proposition *q**. If *q** is true, then a police officer who relies on a driver's race when deciding whom to stop is more likely to stop people who are in fact carrying contraband than he would if he did not rely on it. The law might nonetheless have compelling reasons to prohibit police officers from relying on race when deciding whom to stop. Effective law enforcement is one goal worth pursuing, but not the

a dramatically disproportionate share of street crime in the United States. This is a sociological fact, not a figment of the media's (or the police's) racist imagination.”).

74. See, e.g., Lawrence Blum, *Stereotypes and Stereotyping: A Moral Analysis*, 33 *Phil. Papers* 251, 260 (2004) (“[S]tereotypes are, or involve, not merely generalizations, but false or misleading generalizations, i.e., overgeneralizations.”).

75. See, e.g., Frederick Schauer, *Profiles, Probabilities, and Stereotypes* 179 (2003) (“[P]eople are often inclined to overestimate the proportion of a particularly salient component within a larger population.”); *id.* at 187 (“Because . . . attributes [like race] . . . are ‘visually accessible, culturally meaningful, and interactionally relevant,’ such factors tend to occupy more of the decisionmaking space than their empirical role would support.”); Armour, *supra* note 36, at 791 (“[T]he typical person tends to perceive race as the overriding factor when the supposed assailant is black.”).

76. Moreover, even if the proposition that *q* is true, some (perhaps many) actors in fact believe that *q*, not because they are aware of the relevant statistical studies, but because they believe that most people believe that *q* is true.

only one. Profiles that include race threaten to increase racial stigmatization and the social isolation of the group stigmatized.⁷⁷ How one comes down in the debate over profiling will depend on the size of these effects, and on the moral weight one assigns to them, as well as the moral weight one assigns to the costs and benefits of viable alternatives.⁷⁸

What are the comparable costs and benefits when we turn to self-defense, assuming for the moment that an actor could choose or decide whether or not to form the belief that p based on the belief that q ?⁷⁹ If q is

77. Compare Schauer, *supra* note 75, at 189 (“[U]nder circumstances of existing stigmatization by race or ethnicity for members of certain races or ethnic groups it again might well be worth paying a social price just in order to avoid any further racial or ethnic stigmatization.”), Bernard E. Harcourt, *Rethinking Racial Profiling: A Critique of the Economics, Civil Liberties, and Constitutional Literature, and of Criminal Profiling More Generally*, 71 U. Chi. L. Rev. 1275, 1375–76 (2004) (“[R]acial profiling is an excellent example of how criminal profiling accentuates embedded prejudices in the criminal law.”), and Kennedy, *supra* note 73, at 33 (“[D]efenders of racial profiling frequently neglect the costs of the practice. They unduly minimize (or ignore altogether) the large extent to which racial profiling constantly adds to the sense of resentment felt by blacks of every social stratum toward the law enforcement establishment.”), with Michael Levin, *Responses to Race Differences in Crime*, 23 J. Soc. Phil. 5, 12 (1992) (“[I]f ‘racism’ means unjustified race-consciousness, race-based differentiations need not be racist. In particular, race-based screening is not ‘racist’ if justified by differential crime rates.”).

78. Compare Schauer, *supra* note 75, at 197 (“[E]ven when race is a substantial factor, and thus even when its exclusion would significantly decrease law-enforcement efficiency, the consequences of excluding race from the profile is an increase in crime only if we are holding cost and efficiency constant.”), Samuel R. Gross & Debra Livingston, *Essay, Racial Profiling Under Attack*, 102 Colum. L. Rev. 1413, 1437–38 (2002) (“[W]e should be deeply suspicious of racial profiling, however mild the government’s actions and however justified they may appear.”), and Kennedy, *supra* note 73, at 34 (“Our commitment to a just social order should prompt us to end racial profiling even if the generalizations on which the technique is based are buttressed by empirical evidence.”), with Mathias Risse & Richard Zeckhauser, *Racial Profiling*, 32 Phil. & Pub. Aff. 131, 144 (2004) (“We submit . . . that in a range of plausible cases, utilitarian considerations support racial profiling.”), and Peter H. Schuck, *A Case for Profiling*, *American Lawyer*, Jan. 2002, at 59, 61 (“A wise policy will insist that the justice of profiling depends on a number of variables.”). For a reply to Risse and Zeckhauser, see Annabelle Lever, *Why Racial Profiling Is Hard to Justify: A Response to Risse and Zeckhauser*, 33 Phil. & Pub. Aff. 94 (2005). For thoughts on whether or not the law permits racial profiling in the context of highway drug interdiction, see, e.g., Samuel R. Gross & Katherine Y. Barnes, *Road Work: Racial Profiling and Drug Interdiction on the Highway*, 101 Mich. L. Rev. 651, 744 (2002) (summarizing conclusions based on analysis under the Fourth Amendment and the Equal Protection Clause).

79. This assumption is rejected in part II.

true, then an actor who relies on a putative assailant's race when deciding to shoot or wait is more likely to stop someone who is in fact a deadly aggressor than he would be if he did not rely on it. The law might nonetheless once again have compelling reasons to prohibit an actor's reliance on race. Just as official recognition of q^* may threaten to increase racial stigmatization and the social isolation of the stigmatized group, so too may official recognition of q . How one comes down on the question will once again likely depend on the size of these effects and on the moral weight one assigns to them.⁸⁰

One could of course make the calculus even more complex. For example, one might argue that official stigma is not the only cost involved if the law countenances an actor's belief that q . Corrupting the jury's search for the truth might be another.⁸¹ If an actor claiming self-defense is permitted to introduce evidence at trial designed to substantiate the truth of the proposition that q in order to establish the reasonableness of his belief that p , the process of exposing the jury to such evidence might end up distorting *its* deliberations. In other words, evidence intended to establish the statistical validity of a race-based generalization might end up causing racial stereotypes to taint the jury's verdict. Consequently, white defendants who

80. See, e.g., Randall Kennedy, *Race, Crime, and the Law* 165 (1997) ("Racially discriminatory self-protective action by private persons reinforces existing mistrusts and resentments and circulates them throughout the various spheres of society, public as well as private."); Armour, *supra* note 36, at 795 ("[H]astier use of force against blacks forces blacks who do not want to be mistaken for assailants to avoid ostensibly public places . . . and core community activities. . ."); Kelman, *supra* note 36, at 816 ("[Y]oung black men are stigmatized, excluded from participation in generally available activities . . . subjected to the demeaning supposition that others know a lot about them when who they truly are as individuals is wholly misassessed.").

81. See Armour, *supra* note 36, at 795 ("[R]ace-based evidence of reasonableness impairs the capacity of jurors to rationally and fairly strike the balance between the costs of waiting (increased risk for the person who perceives imminent attack) and the costs of not waiting (injury or death to the immediate victim, exclusion of blacks from core community activities, and, ultimately, reduction of individuals to predictable objects)."). For ideas about how the law might counteract the effects of prejudice on jury decision making, see, e.g., Jody Armour, *Stereotypes and Prejudice: Helping Legal Decisionmakers Break the Prejudice Habit*, 83 Cal. L. Rev. 733, 768 (1995) (Group "references that challenge . . . factfinders to reexamine and resist their discriminatory responses enhance the rationality of the fact-finding process."); Lee, *supra* note 28, at 252–53 (proposing that judges give "race-switching" instructions in appropriate cases).

claim to have killed blacks in self-defense will more often be acquitted on grounds of self-defense than they should be. This potential result is another cost that one must take into account.

For present purposes, I will simply assume that the calculus comes out against Goetz,* and that he should not have possessed the belief that *q*. Thus, Goetz* should not have formed the belief that *p* because he should not have been a racist, i.e., he should not have possessed the belief or desire that *q*. Still, the belief upon which Goetz* “acted” was the belief that *p*, not the belief that *q*, and the desire on which he “acted” was the desire to save his life, not the desire that *q*. He did not “act upon” his racist belief or desire.⁸² Those mental states entered the picture not at the point of action, but earlier, at the point of belief formation.⁸³ Thus, if the law refuses to credit Goetz*’s claim of self-defense, it does so in the end because, according to the character theory, he was a racist. His racism caused him to form the belief that *p*, which caused him to form the belief that he ought to kill his attacker, which caused him to form the intent to kill his attacker, and so forth. Consequently, although Goetz* is convicted of murder, what he did wrong was to be who he was. What he did wrong was to be a racist.

82. When we say that an actor acted with “discriminatory intent,” one thing we might mean is that the belief on which the actor “acts,” though itself unobjectionable, is nonetheless based in part on an objectionable stereotype. The stereotype is a but-for cause of the unobjectionable belief. See David A. Strauss, *Discriminatory Intent and the Taming of Brown*, 56 U. Chi. L. Rev. 935, 956–59 (1989) (proposing this definition of “discriminatory intent”); see also Michael Selmi, *Proving Intentional Discrimination: The Reality of Supreme Court Rhetoric*, 86 Geo. L.J. 279, 289 (1997) (“What the [Supreme] Court means by [discriminatory] intent is that an individual or group was treated differently because of race. . . . [T]he key question is whether race made a difference in the decisionmaking process, a question that targets causation, rather than subjective mental states.”); Amy L. Wax, *Discrimination as Accident*, 74 Ind. L.J. 1129, 1138–39 (1999) (distinguishing between two different meanings of “intentional” as that term might be used in antidiscrimination law, including a “causal view”). The actor may or may not be aware that he possesses the stereotype or that his unobjectionable belief is based on that stereotype. See Strauss, *supra*, at 960 (noting that the causal account of discriminatory intent “reaches both conscious and unconscious discrimination”). On this view Goetz* clearly acted with “discriminatory intent.” Having said that, it is one thing to impose civil liability for acting with such intent. It is another to deny an otherwise-available defense to criminal liability.

83. If one nonetheless insists that Goetz* did “act on” the belief that *q* at the moment he pulled the trigger, then it would seem to follow that he acted on *all* the other background beliefs causing him to form the belief that *p*. But it seems quite implausible to say that whenever we act on a belief we also act on all the background beliefs causing its formation.

Now, Goetz* is not punished *just* because he is a racist. Indeed, one might insist that he is not punished for being a racist at all. On the contrary, he is punished because he intentionally killed someone. True, but he intentionally killed someone only because he believed that he was about to be killed, and he believed that he was about to be killed only because he was a racist who had the misfortune to find himself in a situation in which his racism, combined with other facts about the situation, caused him to form the belief that *p*. So although he is not punished just for being a racist, he is punished because he was a racist unlucky enough to find himself in a situation in which his racism caused him to believe that *p*, and having no reason to ignore that belief, he formed the intention to kill and then executed that intention in action.

B. Punishment for Being a Racist

No state can coherently punish a person for something unless the person punished is responsible for it. Was Goetz* responsible for the racist beliefs or desires he possessed, which beliefs and desires caused him to form the belief that *p*, and ultimately to pull the trigger? According to the character theory, he was.

The character theory says that we are responsible for the standing beliefs and desires constitutive of our characters, such as Goetz*'s standing belief and desire that *q*, because and just because *we are* our characters.⁸⁴ In other words, we are responsible for who we are just because we are who we are. Indeed, our common practices of praise and blame presuppose some such responsibility. We praise people for their virtues and blame them for their vices. Our praise and blame are often directed at the person for *being* this or that, and not just at what he has *done* or *failed to do* because he is this or that. Consequently, we are free to, and indeed should, blame Goetz* for possessing the racist beliefs and desires making him a racist, and we are free to do so without regard to how or why he came to possess them. Simple possession is enough.

Yet the question is not whether *we* can condemn Goetz* for being a racist. The question is whether the *state* can, and more precisely, whether

84. See, e.g., Moore, *supra* note 18, at 571 (“[W]e are responsible for our character because we are, in part, constituted by our characters.”).

the state can condemn him for being a racist through the hardship that makes *punishment* what it is. A liberal state need not treat the racists in its midst the same as it does those who accord their fellow citizens the concern and respect to which they are due. The state is free to criticize them for their racist characters. It might also refuse to do business with them. The KKK Construction Co. should not be disappointed if the state decides to contract with another firm. Nonetheless, any recognizably liberal state has no authority to punish its citizens for who they are, no matter what the content of their character.⁸⁵ Nor should the content of an actor's character form the basis upon which the state denies him a defense to which he would otherwise have been entitled. If a liberal state cannot make it a crime to be a racist, then neither should it permit only non-racists to gain access to otherwise available defenses, and deny access to racists.

The analogy here would be to so-called status offenses. Our law does not in fact punish actors just for being who they are,⁸⁶ nor should it. It might punish them for doing things that result in them being who they are, or even for not doing things to try to change who they are. But it does not, nor should it, punish them just for who they are. Again, the reason is not that we are not responsible for who we are. We are responsible for who we are, even if we have not chosen to be who we are, just because we are

85. See, e.g., George Sher, In Praise of Blame 69 (2006) (arguing that it is permissible to blame a person for his character, even if he is not responsible for it, but not to punish him for it); Robert Merrihew Adams, Involuntary Sins, 94 *Phil. Rev.* 3, 21 (1985) (arguing that it is permissible to hold responsible and to blame a person for his character, but not to punish him for it); Angela M. Smith, Responsibility for Attitudes: Activity and Passivity in Mental Life, 115 *Ethics* 236, 270–71 (2005) (arguing that we are responsible for our beliefs but also noting that “[o]ne question . . . is whether we are open to the very same *kinds* of appraisals for our [beliefs] as we are for our voluntary actions”) (emphasis added). But cf. Tadros, *supra* note 40, at 263 (stating that a “defendant who forms a false belief about the risks in a particular case” based on “prejudiced background beliefs” is an “exception” to the “general principle” that defendants who unwittingly impose risks do “not show the appropriate kind and degree of fault required for the proper imposition of criminal responsibility”); Andrew E. Taslitz, Condemning the Racist Personality: Why the Critics of Hate Crime Legislation Are Wrong, 40 *B.C. L. Rev.* 739, 742 (1999) (arguing that a “vision of virtuous citizen character in a republic . . . requires us to condemn [and punish] the racist personality”).

86. See, e.g., *Robinson v. California*, 370 U.S. 660, 666–67 (1962) (holding that a “state law which imprisons a person” for the “‘status’ of narcotics addiction” violates the Eighth Amendment because it “inflicts a cruel and unusual punishment”).

who we are. The reason is that the responsibility we bear for our characters in virtue of the fact that we are our characters, though strong enough to underwrite some forms of blame and censure, is not strong enough to underwrite state punishment.

Proposed amendments to the character theory do nothing to remedy this problem. For example, one might argue that an actor is responsible for his character and thus liable to punishment for his character, *unless* he lacked the capacity or a fair opportunity to choose or otherwise shape his character, such that his character is not his own, in which case he is neither responsible nor liable to punishment. Or one might argue that such an actor, though he remains responsible for who he is and thus liable to punishment, should nonetheless receive the state's mercy, or at least be a candidate for its mercy.⁸⁷ For example, suppose that Goetz* possessed his racist belief or desire only because he was the victim of prior attacks involving black assailants.⁸⁸ Indeed, suppose he despises himself for what he has become, and has tried without success to rid himself of his racism. Or suppose Goetz* had been brainwashed into being a racist. Under the proposed amendments to the character theory, Goetz might not forfeit his claim to self-defense, or he might forfeit his claim but nonetheless catch a break in the name of mercy.

These amendments improve the character theory, but they hardly fix it. Why not? Because if and when an actor *is* punished, either because he has no excuse for his character or because he is denied mercy, it remains the case that he is being punished, not for anything he has done or failed to do, but for who he is.⁸⁹ The target of the state's punishment continues to

87. See, e.g., Dan M. Kahan & Martha C. Nussbaum, Two Conceptions of Emotion in Criminal Law, 96 Colum. L. Rev. 269, 366–72 (1996) (making this suggestion).

88. See, e.g., Armour, *supra* note 36, at 799 (describing such a person as an “[i]nvoluntary [n]egrophobe”). Armour argues that such an actor should not be permitted to claim self-defense because “[l]egal recognition of the Involuntary Negrophobe's claims would subvert the general welfare by destroying the legitimacy of the courts.” *Id.* at 802.

89. See, e.g., Moore, *supra* note 18, at 585 (“That punishment would be deserved because of bad character alone is something the character theorist seems committed to, however much other values prevent punishment of this class of deserving persons.”). One might of course take the view that no one is ever responsible for his character, in which case one would end up being a skeptic about the possibility of moral responsibility altogether. See, e.g., Robert Kane, *A Contemporary Introduction to Free Will* 121 (2005) (describing the thesis that free will requires ultimate responsibility which in turn requires that “we must be responsible for forming the wills or characters that now determine our acts”).

be his character, but a liberal state worthy of the name cannot take character to be the ultimate target of state punishment. Thus, while Goetz* is no doubt responsible for his character, he cannot be punished for it, nor therefore can he be punished for it through the backdoor workings of a forfeiture rule.

III. THE CHOICE THEORIES

The character theory of unreasonableness says that Goetz*'s belief that he was about to be killed was unreasonable, such that he forfeits his claim to self-defense, because he was a racist. The forfeiture is based on the content of his character. In contrast, the remaining two theories base the forfeiture on some choice he made but should not have made, or on some choice he failed to make but should have made. These theories therefore ground Goetz*'s responsibility for believing that *p* in some choice he made. While a liberal state cannot legitimately punish a person for who he is, it can legitimately punish him for the choices he makes (or at least for some of those choices). Likewise, while a liberal state cannot legitimately deny an actor a defense based on who he is, it can deny him a defense based on the choices he makes (or at least some of those choices).

We can distinguish two choice-based theories of unreasonableness. According to the *belief-choice theory*, Goetz*'s belief that *p* was unreasonable because he chose to believe that *p* when he should have chosen not to believe that *p*. The object of the forbidden choice is the belief that *p*. According to the *character-choice theory*, Goetz*'s belief that *p* was unreasonable because he should not have chosen to possess the racist beliefs or desires causing him to form the belief that *p*, but he did so choose; or he should have chosen to dispossess himself of them, but he failed to do so. In short, he chose to become or remain a racist when he should not have so chosen. The objects of the forbidden choice are those acts or omissions that caused him to possess the belief or desire that *q*.⁹⁰

90. According to another choice-based theory (not discussed in the text), Goetz* should lose his claim of self-defense, not because he chose to believe that *p*, nor because he chose to be a racist, but because he failed to stop his racist beliefs from causing him to form the belief that *p*: He failed to exercise doxastic-self control when he could and should have exercised such self-control. In other words, he should have stopped his stereotypical beliefs

A. The Belief-Choice Theory

Recall Goetz*'s psychology at the moment he pulled the trigger.⁹¹ Starting with his network of background beliefs and desires, including the belief or desire that *q*, Goetz* may have thought about whether his life was in immediate danger, or he may not have. Either way, he formed the belief that it

from being activated in the first place, or if he failed at that, he should have stopped his activated stereotypical beliefs from causing him to form the belief that *p*. If such self-control is possible, its exercise is unlikely to be subject to one's conscious will. See, e.g., John A. Bargh, *The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effects*, in *Dual-Process Theories in Social Psychology* 361, 378 (Shelly Chaiken & Yaacov Trope eds., 1999) ("[T]he evidence to date concerning people's realistic chances of [consciously] controlling the influence of their automatically activated stereotypes weighs in heavily on the negative side."); Timothy D. Wilson et al., *Mental Contamination and the Debiasing Problem*, in *Heuristics and Biases: The Psychology of Intuitive Judgment* 185, 200 (Thomas Gilovich et al. eds., 2002) (expressing "pessimis[m] about people's natural ability to willfully control and correct their [contaminated] judgments" though "by no means suggesting that reducing mental contamination is a lost cause"). But see Nilanjana Dasgupta & Luis M. Rivera, *From Automatic Antigay Prejudice to Behavior: The Moderating Role of Conscious Beliefs about Gender and Behavioral Control*, 91 *J. Personality & Soc. Psych.* 268, 277 (2006) ("[T]he present data illustrate that relatively spontaneous interpersonal actions can be modified by motivation and control[. F]uture research might investigate whether . . . [these results] generalize to other . . . actions and decisions that are more constrained by cognitive load or time pressure."); Patricia G. Devine & Margo J. Monteith, *Automaticity and Control in Stereotyping*, in *Dual-Process Theories in Social Psychology*, supra, at 339, 355 (discussing "findings [that] provide reason for optimism that control over stereotyping is possible").

Instead, the self-control needed to counteract the automatic influence of stereotypes on belief formation is probably best portrayed as a sophisticated mental habit operating in much the same unconscious and automatic manner as the stereotypes it fights. The idea is to enlist a good habit to neutralize a bad one. See, e.g., Patricia G. Devine et al., *Breaking the Prejudice Habit: Progress and Obstacles*, in *Reducing Prejudice and Discrimination* 185, 202 (Stuart Oskamp ed., 2000) ("For low-prejudice people who already possess the requisite internal motivation to overcome prejudice, the challenge is to learn the skills necessary to respond consistently with their nonprejudiced beliefs."); John F. Dovidio et al., *Reducing Contemporary Prejudice: Combating Explicit and Implicit Bias at the Individual and Intergroup Level*, in *Reducing Prejudice and Discrimination*, supra, at 137, 145 ("[S]elf-regulation, extended over time, may produce changes even in previously automatic, implicit negative responses."); Jack Glaser & John F. Kihlstrom, *Compensatory Automaticity: Unconscious Volition Is Not an Oxymoron*, in *The New Unconscious* 171, 171 (Ran R. Hassin et al. eds., 2005) ("[U]nconscious vigilance for bias can lead to corrective processes that also operate without awareness or intent."); Margo J. Monteith et al., *Putting the Brakes on Prejudice: On the Development and Operation of Cues for Control*, 83 *J. Personality & Soc. Psychol.* 1029, 1045 (2002) ("[P]eople can learn to put the brakes

was, and that killing was necessary to prevent being killed. Likewise, he may have thought about what he ought to do, or he may not have. Again, either way, he formed the twin beliefs that he ought to save himself and that he ought to kill his assailant in order to accomplish that end. He then chose to kill, causing himself to form the intent to kill, and finally, he chose to execute that intent in action, causing his finger to pull the trigger.

Goetz* is in control at four points in this sequence. First, he is in control if and when he thinks about whether or not his life is in danger. Thinking, deliberating, reflecting and so forth are mental acts over which we have some measure of control.⁹² Second, he is in control if and when he thinks about what he ought to do. Again, thinking, deliberating, reflecting and so forth are mental acts over which we have some measure of control. Third, he is in

on their prejudices and control the influence of processes that otherwise could result in racially biased behavior"); Kerry Kawakami et al., *Just Say No (to Stereotyping): Effects of Training in the Negation of Stereotypic Associations on Stereotype Activation*, 78 *J. Personality & Soc. Psychol.* 871, 884 (2000) ("[P]articipants who received extensive training in negating stereotypes were able to reduce . . . stereotype activation."); Gordon B. Moskowitz et al., *Preconsciously Controlling Stereotyping: Implicitly Activated Egalitarian Goals Prevent the Activation of Stereotypes*, 18 *Soc. Cognition* 151, 173 (2000) ("[C]hronic [egalitarian] goals disrupt stereotype activation.").

This alternative theory is perhaps best understood as a variation on the character-choice theory. See *infra* pp. 162–70. The character-choice theory says that Goetz* loses his defense if and because he chose to become or remain a racist. The alternative theory says that he loses his defense if and because he chose not to develop, or at least chose not to try to develop, the right cognitive habits. See, e.g., Simon Wigley, *Automaticity, Consciousness and Moral Responsibility*, 20 *Phil. Psych.* 209, 218 (2007) ("[A]n automatic agent is praiseworthy or blameworthy not because of the immediate actions that led to a good or bad outcome, but rather because of what they did, or omitted to do, in the past."). Accordingly, it might be called the habit-choice theory. Could a liberal state make it a crime for a citizen to fail to try to develop such a habit? For example, could a liberal state make it a crime for a citizen to fail to attend a diversity training program the goal of which is to instill the requisite habit of doxastic self-control? If not, then neither should it be permitted to base the forfeiture of an otherwise valid claim of self-defense upon such an omission. In any event, it bears noting that the prejudice habit is apparently easier to acquire than it is to break. See Aiden P. Gregg et al., *Easier Done Than Undone: Asymmetry in the Malleability of Implicit Preferences*, 90 *J. Personality & Soc. Psychol.* 1, 17 (2006) ("[P]eople can speedily develop, at an implicit level, unfavorable and undeserved evaluations of social groups that they can only laboriously unburden themselves of them later.").

91. See *supra* text accompanying note 27.

92. See, e.g., Nomy Arpaly, *Merit, Meaning, and Human Bondage: An Essay on Free Will* 96 (2006) ("[R]eflection, like fishing or fact finding, is a process we can decide to initiate but whose results we cannot choose."). It might be more accurate to say that we can

control when he chooses to form the intent to kill. Choosing, like thinking, is a mental act over which we have control.⁹³ Fourth, he is in control when he chooses to execute that intention, transforming it into action. Again, choosing is a mental act over which we have control. At each of these moments, Goetz* has *done* something that he might not have done, even if that which he has done is a mental act, and not a bodily one.

More important is when he is *not* in control. He is not in control at the moment he forms the belief that his life is in danger, and that killing was the only way to save himself. He has no control over whether he forms that belief at that moment or not. He cannot will, decide, or choose to believe that *p* or not-*p*.⁹⁴ None of us has such direct control over our beliefs. We have direct control over our choices and actions, but not over our beliefs. The only control we have over our beliefs is indirect. We can act or fail to act in ways that affect the evidence available to us, which can in turn affect the beliefs we form. We can also act or fail to act in ways that affect our cognitive capacities and habits, which can in turn affect the beliefs we form. We can also reflect on the beliefs we have formed, after we have formed them, and we can choose whether or not to endorse or disavow those beliefs.⁹⁵ Yet whether we possess them or not in the first place is not up to us. Our beliefs just happen to us when they happen.

choose to think, but we cannot choose not to think, though we can choose to do things to try to distract ourselves from thinking.

93. See, e.g., Robert Kane, *The Significance of Free Will* 24 (1996) (“Choices and decisions are acts of mind (or will), and hence events that happen at a time, possibly terminating deliberations and giving rise to intentions.”); Alfred R. Mele, *Motivation and Agency* 210 (2003) (“[P]ractical deciding [i]s a momentary mental action of intention formation.”).

94. See, e.g., David Owens, *Reason Without Freedom: The Problem of Epistemic Normativity* 85 (2000) (“[B]elief is not subject to the will.”); William P. Alston, *The Deontological Conception of Epistemic Justification*, 2 *Phil. Perspectives* 257, 263 (1988) (“[W]e are not so constituted as to be able to take up propositional attitudes at will.”); Neil Levy, *Doxastic Responsibility*, 155 *Synthese* 127, 148 (2007) (concluding among other things that “arguments purporting to establish that we have direct control over our beliefs are not persuasive”); Dion Scott-Kakures, *On Belief and Captivity of the Will*, 54 *Phil. & Phenomenological Res.* 77, 77 (1994) (arguing that it is conceptually, and not merely contingently, true that “with respect to our beliefs our wills are captive”); Bernard Williams, *Deciding to Believe*, in Bernard Williams, *Problems of the Self* 136, 148 (1973) (“[I]t is not [merely] a contingent fact that I cannot bring it about, just like that, that I believe something.”).

95. See, e.g., L. Jonathan Cohen, *An Essay on Belief and Acceptance* 22 (1992) (“Acceptance, in contrast with belief, occurs at will. . . .”); Stephen Shute, *Knowledge and*

If so, then the belief-choice theory is a nonstarter. It begins from a false premise. It presupposes that Goetz* should not have chosen to form the belief that *p* because that belief was based in part on racist beliefs or desires, and that Goetz* should therefore forfeit his claim to self-defense because, contrary to what he should have done, he chose to believe that *p*. But Goetz* did not choose to believe that *p*. He did not choose to believe that *p* because he could not have so chosen. He may still be responsible for forming the belief that *p*,⁹⁶ just as he is responsible for his

Belief in Criminal Law, in *Criminal Law Theory*, supra note 63, at 171, 192 (“Acceptances . . . engage the will in a different way [than do beliefs]. Beliefs are ‘passive.’ They cannot be acquired directly through an act of will. . . . In contrast, acceptances are ‘active’; they do respond to will.”). But cf. Raimo Tuomela, *Belief Versus Acceptance*, 2 *Phil. Explorations* 122, 136 (2000) (concluding that “acceptance need not be intentional action, [and thus] the differences between belief and acceptance do not boil down to the simple view that acceptance, contrary to belief, is based on the agent’s direct exercise of his will”).

96. Some writers argue that we have the same sort of control over our beliefs as we do over our actions, and as such, that we bear the same responsibility for our beliefs as we do for our actions. See, e.g., Carl Ginet, *Deciding to Believe*, in *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue* 63, 63 (Matthias Steup ed., 2001) (defending the “naïve intuition that coming to believe something just by deciding to do so is possible”); Keith Frankish, *Deciding to Believe Again*, 116 *Mind* 523, 523 (2007) (defending the “view that we can form beliefs directly”); Christoph Jäger, *Epistemic Deontology, Doxastic Voluntarism, and the Principle of Alternative Possibilities*, in *Knowledge and Belief* 217, 226 (Winfried Löffler & Paul Weingartner eds., 2004) (concluding that “there is a crucial sense in which we hold [beliefs] freely” and that this suffices for holding us responsible for our beliefs); Sharon Ryan, *Doxastic Compatibilism and the Ethics of Belief*, 114 *Phil. Stud.* 47, 70 (2003) (“If you have compatibilist intuitions, you should deny [the] premise [that doxastic attitudes are never under our voluntary control].”); Matthias Steup, *Doxastic Freedom*, 161 *Synthese* 375, 375 (2008) (“Compatibilism entails that our actions and our doxastic attitudes [including our beliefs] are mostly free.”). But see Nikolaj Nottelmann, *The Analogy Argument for Doxastic Voluntarism*, 131 *Phil. Stud.* 559, 559 (2006) (rejecting arguments that “belief formations may qualify as voluntary in perfect analogy to certain types of actions or even to actions in general”).

Other writers argue that we do not have the same control over our beliefs as we do over our actions, but that we bear some responsibility for our beliefs nonetheless. See, e.g., Adams, supra note 85, at 17 (“[B]lameworthiness of states of mind[, including beliefs,] is not dependent upon voluntariness.”); Robert Audi, *Doxastic Voluntarism and the Ethics of Belief*, in *Knowledge, Truth, and Duty*, supra, at 93, 105 (The conclusion that “neither believing nor forming beliefs is a case of action” does not “prevent our sustaining a deontic version of an ethics of belief”); Richard Feldman, *Voluntary Belief and Epistemic Evaluation*, in *Knowledge, Truth, and Duty*, supra, at 77, 90 (concluding that “deontological

character,⁹⁷ but whatever responsibility he bears for that belief is too weak to support punishing him for it. A liberal state cannot punish an actor for a choice he never made, nor can a choice he never made be the basis for denying him a defense to which he would otherwise have been entitled.

B. The Character-Choice Theory

The second choice-based theory of unreasonableness, unlike the first, does manage to get off the ground. It says that Goetz*'s belief that *p* was unreasonable, not because he should not have chosen to believe that *p* (but did so choose), but rather because he should not have chosen to possess the racist beliefs or desires causing him to form the belief that *p* (but did so choose). Because he chose to possess those beliefs or desires, and because he should not have so chosen, he loses his claim of self-defense if and when those beliefs or desires cause him to form the belief that *p*. In other words, Goetz* loses his claim of self-defense, not just because he is a racist, but because he chose to become or remain a racist.

The analogy here is to crimes of possession. The law does not punish an actor just because he possesses an item the law does not permit him to possess. Instead, it punishes him if and because he has done something to come into possession of it, realizing what it is that he has come to possess; or if and because, upon realizing that he is already in possession of the prohibited item, he fails to do something to dispossess himself of it.⁹⁸

judgments about belief . . . do not imply that belief is voluntary”); Pamela Hieronymi, *Responsibility for Believing*, 161 *Synthese* 357, 358 (2008) (“[O]n at least one plausible account of what it is for a thing to be voluntary and what it is to be responsible for something, beliefs are not voluntary and yet, for failing to be voluntary, they are a central examples of the sort of thing for which we are most fundamentally responsible.”); Nishi Shah, *Clearing Space for Doxastic Voluntarism*, 85 *Monist* 436, 436 (2002) (“While I agree . . . that agents don’t have the capacity to decide what to believe, I disagree that the application of deontological concepts requires this kind of control.”); Smith, *supra* note 85, at 271 (“[W]hat makes us responsible for our attitudes[, including our beliefs], is not that we have voluntarily chosen them . . . but that they are the kinds of states that reflect and are in principle sensitive to our rational judgments.”).

97. See *supra* part II.B.

98. See Model Penal Code § 2.01(4) (Proposed Official Draft 1962) (“Possession is an act . . . if the possessor knowingly procured or received the thing possessed or was aware of his control thereof for a sufficient period to have been able to terminate his possession.”).

The target of the state's punishment is not possession of the proscribed item itself. The target is the actor's choice to cause himself to come into possession of it, or the choice to retain possession of it when he could and should have gotten rid of it.

C. Punishment for Choosing to Be Racist

Basing a forfeiture rule upon a choice an actor should not have made avoids the problem associated with the character-based theory of unreasonableness, inasmuch as the forfeiture is triggered not by the actor's character, but rather by a choice he makes. Likewise, it avoids the problem associated with the belief-choice theory, inasmuch as that which triggers the forfeiture is a choice over which the actor has control, not a choice over which he has no control, which is to say no choice at all. Nonetheless, the character-choice theory has at least three problems of its own, which can be grouped under the headings of luck, legality, and liberalism.⁹⁹

1. Luck

The first problem with the character-choice theory is luck. Suppose that Goetz* was offered a substantial sum of money to try to become a racist. Wishing to collect, he decided to join the Ku Klux Klan, hoping or believing that he would thereby become a racist. Assume that as a result of that choice he does indeed succeed in transforming himself into a racist. Now assume (after collecting his cash) that he is unlucky enough to find himself in a situation in which, because he is a racist, he forms the mistaken belief that a black person is about to kill him. If as a result of that mistaken belief he kills his imagined assailant, he will be guilty of murder—or at least he will be if his belief that *p* is unreasonable, which we are assuming it is, inasmuch as he would not have formed that belief but for his earlier choice to join the KKK.

Now consider Goetz**. He too joins the KKK, hoping or believing that he will thereby turn himself into a racist. Unlike Goetz*, Goetz**'s

99. In addition to the problems grouped under these headings is the problem of disproportionality associated with forfeiture rules in general.

efforts fail. Despite his regular attendance at rallies, cross burnings and the like, he never ends up believing racist stereotypes or harboring racial animus. He is not even indifferent. Or suppose that he does succeed in turning himself into a racist, but that he, unlike Goetz*, is lucky enough never to find himself in a situation in which his racism causes him to believe that he is about to be killed. Goetz* and Goetz** made the same choice. Each chose to join the KKK hoping or believing that he would thereby become a racist. Likewise, neither chose to place himself in a situation in which he realized that his racism would or might cause him to believe that *p*.¹⁰⁰ The only thing setting them apart is luck. Bad luck for Goetz*. Good luck for Goetz**. Goetz* is guilty of murder. Goetz** is guilty of nothing.

Perhaps that outcome should not be disturbing. Rightly or wrongly, criminal liability often depends on all kinds of luck.¹⁰¹ For example, murderers are punished more severely than attempted murderers, even if luck is the only thing that sets them apart. If you sneeze just as the trigger is pulled, thereby missing the target, then the crime is attempted murder. If you do not sneeze, and the target is killed, then the crime is murder. Perhaps the desire to purge all luck from the criminal law is a desire destined never to be fulfilled. Perhaps we shouldn't worry about luck's influence on the fates of Goetz* and Goetz**. Perhaps. Yet even if their different fates are not a problem, or not much of one, the character-choice theory has two more.

100. One could say that many crimes happen just because the people who commit them find themselves in unlucky situations amenable to their commission. For example, an actor who finds himself alone with an unattended cash register and who as a result decides then and there to commit larceny was, one could say, unlucky enough to find himself alone with the cash register. But we hardly feel any sympathy for the unlucky larcenist. If so, then why should we feel any sympathy for the unlucky Goetz* who likewise finds himself in an unlucky situation? One salient difference between the unlucky larcenist and the unlucky Goetz* is that the former, but not the latter, chooses to do that which he presumably believes he is not permitted to do. The unlucky larcenist is presumably aware of the fact that he is committing a crime. In contrast, the unlucky Goetz* presumably believes under the circumstances as he believes them to be that he is permitted to kill.

101. See, e.g., Moore, *supra* note 18, at 235 (distinguishing between result luck, luck in execution, planning luck, and constitutive luck).

2. Legality

The second problem with the character-choice theory is legality. Forfeiture rules covertly criminalize the conduct upon which the forfeiture is based. For example, if an actor chooses to place himself in a situation in which he will or might be subject to a threat, and the threat materializes, he loses any defense to which he might otherwise have been entitled if he commits a crime in order to avoid the threat. Of course, an actor who makes such a choice will not be punished unless and until the threat materializes; nor will he be punished unless and until he commits a crime in response to it. In the law's eyes, moreover, he is guilty of the crime committed in response to the threat, not for choice to expose himself to the threat in the first place. But the law blinds itself to reality here, since the only thing the actor has chosen to do that he should not have done is to risk exposing himself to a threat he ought instead to have chosen to avoid.

So far as I know, no state makes it a crime to choose to expose oneself to a threat. But a state could criminalize that choice if wanted to. A state could, if it wanted, make it a free-standing crime to choose to place oneself in a threatening situation, whether or not the anticipated threat comes to pass, and whether or not the actor commits a crime in an effort to avoid it if it does come to pass.¹⁰² The same goes for the intoxication forfeiture rule. A state could, if it wanted, make it a crime to consume intoxicating substances, no matter what happens thereafter. The wisdom of such a crime might be open to question, but the fairness of enforcing it would not be, assuming one was aware of the new prohibition. Compliance would not be difficult. Don't drink. Avoid situations that you realize might be threatening. That's all it would take.

But now imagine that a legislature adds the following provision to its penal code: Whoever believes unreasonably shall be guilty of a felony. Now we have a problem. Wouldn't such a provision be unduly vague, leaving those subject to it without fair notice as to how to comply? Are you believing unreasonably now?

Indeed, matters are even worse. A statute making it a crime to believe unreasonably would, because it is so vague, constitute a delegation of

¹⁰². See, e.g., Sangero, *supra* note 36, at 337–39 (discussing a possibility along these lines).

law-making power to prosecutors, jurors, and judges.¹⁰³ Prosecutors would decide when a person has believed unreasonably when they decide to bring a prosecution. Juries would decide when a person has believed unreasonably when they decide to return a conviction, and judges would decide when a person has believed unreasonably when they decide to uphold a conviction on appeal. Under the reasonable-belief rule of self-defense, an actor commits a crime—enforced via a forfeiture rule—when prosecutors, juries, and judges say that he has committed a crime, and not before. Prosecutors, juries and judges can of course exercise this power only when an actor kills someone. But this limitation on the delegation of the power to define crimes does nothing to legitimize its exercise within the scope of that delegation.

Perhaps the legislature could enact a more specific provision meant to address particular instances of unreasonable believing. Go back to the Goetz* case and the problem of the racism. Perhaps the legislature could make it a crime to choose to become a racist, as the character-choice theory maintains. You commit this crime if you set out on a course of action, either with the goal of becoming a racist or foreseeing that you will or might become one, whether or not you actually do. Or maybe the crime can be made even more specific. Suppose you commit a crime if you join the KKK or associate with skinheads with the purpose of becoming a racist, or foreseeing that you will or might become one. Such a crime would not be unduly vague. Citizens would be able to comply. Just stay away from the KKK, and don't consort with skinheads.

Of course, most of us don't need to join the KKK or hobnob with skinheads in order to acquire beliefs that can fairly be characterized as racist (though such associations might be needed to acquire racist desires). On the contrary, the prevailing wisdom among cognitive scientists is that more or less all of us are burdened with racist beliefs. Some of us are aware of our affliction. We have taken the on-line Implicit Association Test and realize that we automatically associate black with bad.¹⁰⁴ The rest of us are

103. See, e.g., Dan M. Kahan, *Is Chevron Relevant to Federal Criminal Law?*, 110 Harv. L. Rev. 469, 475 (1996) (“[R]esort[] to general statutory language . . . necessarily transfers lawmaking responsibility to courts (or prosecutors).”).

104. See Project Implicit, <https://implicit.harvard.edu/implicit> (last visited Jan. 4, 2008). For the initial research “apprais[ing] the IAT method’s usefulness for measuring evaluative associations that underlie implicit attitudes,” see Anthony G. Greenwald et al.,

strangers to ourselves, blissfully unaware, willfully ignorant, or self-deceived. In addition, most of us manage to become racists without even trying. We don't need to *do* anything to acquire racist beliefs. Though not born with racist beliefs, we soon enough pick them up. For the unlucky, their parents, families, and friends are their first teachers. For the lucky, having grown up among a more enlightened circle of intimates, popular culture steps in to teach the association.¹⁰⁵ Falling into the racism habit is easy.

Perhaps the legislature should therefore make it a crime to fail to try to purge oneself of racist beliefs. If, as it should, this crime required the state to prove that the actor realized he possessed such beliefs, those who in fact possessed such beliefs would not be guilty if those beliefs were tucked away in the unconscious. Given the subtle nature of modern-day racism, this description probably applies to many, if not most, people who possess such beliefs. On the other hand, if the crime did not require the state to prove that the actor realized he possessed such beliefs, the unconscious racist would not escape punishment, but giving him his just deserts would be unjust. Punishing an actor for failing to discharge an obligation is unfair if he is unaware of the facts placing him under that obligation in the first place,¹⁰⁶ even if he need not be aware of the obligation itself.¹⁰⁷

Measuring Individual Differences in Implicit Cognition: The Implicit Association Test, 74 *J. Personality & Soc. Psychol.* 1464, 1464 (1998). For a recent update and “assessment on [the] current status” of the IAT, see Brian A. Nosek et al., *The Implicit Association Test at Age 7: A Methodological and Conceptual Review*, in *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* 265, 266 (John A. Bargh ed., 2007). But see Hal R. Arkes & Philip E. Tetlock, *Attributions of Implicit Prejudice, or “Would Jesse Jackson ‘Fail’ the Implicit Association Test?”*, 15 *Psychol. Inquiry* 257, 257 (2004) (offering “three objections to the inferential leap from the comparative [reaction time] of different associations to the attribution of implicit prejudice”).

105. See, e.g., Jerry Kang, *Trojan Horses of Race*, 118 *Harv. L. Rev.* 1489, 1556 (2005) (“[V]iolent crime stories [on the local news] can . . . exacerbate implicit bias. . .”).

106. See, e.g., Larry Alexander, *Criminal Liability for Omissions: An Inventory of Issues*, in *Criminal Law Theory*, supra note 63, at 121, 124 (noting that “[w]hat authorities there are on [the] point generally agree that liability for failing [to discharge a duty to act] does not attach to those who are unaware of the facts that give rise to the duty”).

107. See *id.* (noting that an actor can be held liable for an omission even though he is unaware of the obligation to act but suggesting that this result might violate the principle of legality).

Maybe the legislature should instead demand that all adults periodically attend state-sponsored diversity training classes, or something along similar lines, with the purpose of divesting its citizens of their racism. Failure to attend would result in a fine. Three or more such failures could mean jail time. Citizens subject to such an obligation would have little trouble meeting it, assuming they were aware of it. All you would need to do is attend class. Consequently, the principle of legality would not stand in the way of the state creating such a crime. Neither would it stand in the way of the state's reliance on an analogous forfeiture rule. But it seems to me that another principle would stand in the way.

3. Liberalism

The third problem with the character-choice theory cuts to the chase. According to the prevailing orthodoxy,¹⁰⁸ a liberal state can legitimately criminalize acts if and because those acts cause or risk causing harm, even if the harm targeted is quite remote,¹⁰⁹ but no other reason will underwrite criminalization. According to J.S. Mill's influential formulation: "[T]he only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others."¹¹⁰ Whatever else the harm principle means, surely it means that a liberal state

108. For proposed replacements to the harm principle, see, for example, Meir Dan-Cohen, *Defending Dignity*, in Meir Dan-Cohen, *Harmful Thoughts: Essays on Law, Self, and Morality* 150, 150 (2002) (arguing that liberalism's "harm principle" is not a "neutral standard" and considering its "replacement . . . by . . . the dignity principle: the view that the main goal of the criminal law is to defend the unique moral worth of every human being"); Arthur Ripstein, *Beyond the Harm Principle*, 34 *Phil. & Pub. Aff.* 215, 215 (2006) (arguing that a "commitment to individual sovereignty within a sphere of action in which you are answerable only to yourself requires that we abandon the harm principle" in favor of "the sovereignty principle").

109. See, e.g., Andrew von Hirsch, *Extending the Harm Principle: "Remote" Harms and Fair Imputation*, in *Harm and Culpability*, *supra* note 41, at 259, 276 (arguing that it is "important to develop fair-imputation principles when dealing with remote risks," lest the harm principle lose its effectiveness as a limit on the state's power to punish).

110. John Stuart Mill, *On Liberty* 13 (Currin V. Shields ed., 1956) (1859). For the classic modern statements of the harm principle, see 1 Joel Feinberg, *The Moral Limits of the Criminal Law: Harm to Others* (1984); H.L.A. Hart, *Law, Liberty and Morality* (1963). The harm principle should be understood as a necessary, but not a sufficient, condition for criminalization in a liberal state. See Douglas Husak, *The Criminal Law As Last Resort*, 24

cannot legitimately criminalize the simple possession of certain beliefs or desires.¹¹¹ Presumably, a liberal state likewise lacks the authority to force a person on pain of punishment to do or not to do things with the purpose of preventing them from possessing state-disapproved beliefs or desires, or with the purpose of causing them to shed such beliefs or desires—even if the state’s further purpose is to prevent the harm that results from Goetz*-like cases.

No one doubts that Goetz* acts with the intent to cause harm when he pulls the trigger on the pistol: He intends to kill. But insofar as Goetz* would—save for the reasonable-belief rule—be entitled to claim self-defense, the harm he intends to cause is one the state itself would permit him to cause. Like anyone else, Goetz* is permitted to kill if and when he believes that he is about to be killed, and that killing is the only way to avoid being killed. The real question is whether a liberal state can deny him that defense because he chose to become or remain a racist. If it can, then presumably it could also punish a citizen whenever he chooses to act or not act in ways likely to cause him to become or remain a racist, where the state’s purpose is to prevent him from becoming a racist, or to transform him into a non-racist. But it is hard to see how a liberal state worthy of the name could do such a thing. On the contrary, one imagines that citizens of liberal states must be free to choose to become or remain racists, and the liberal state has no choice but to tolerate such choices.

Of course, a liberal state is not totally without tools to combat racism. For example, it can coerce children to attend school, and while there to inculcate, or if you prefer, indoctrinate them in the virtues of liberal citizenship, virtues which have no place for the vice of racism. It can likewise try to persuade its adult citizens to reject racism, speaking out loud and clear against it. It might even be free to condition the availability of certain benefits—such as the privileges of carrying a firearm or state employment—on a citizen’s willingness to take part in programs designed to rid participants

Oxford J. Legal Stud. 207, 213–14 (2004). For an argument to the effect that “[c]laims of harm have become so pervasive that the harm principle has become meaningless,” see Bernard E. Harcourt, *The Collapse of the Harm Principle*, 90 J. Crim. L. & Criminology 109, 113 (1999).

111. See, e.g., Shlomit Wallerstein, *Criminalizing Remote Harm and the Case of Anti-Democratic Activity*, 28 Cardozo L. Rev. 2697, 2700 (2007) (“Whatever the exact meaning of the harm principle is, it is indisputable that thoughts and beliefs are excluded from consideration and, therefore cannot be restricted.”).

of any racist beliefs or desires they might possess. Indeed, more controversially, it might even be free to do things *to* its citizens—like expose them to various debiasing stimuli¹¹²—in order to break the implicit association most of them are apt to make between black and bad.

But again, what a liberal state cannot do is to force its citizens on pain of punishment to do or not do things with the purpose of ridding them of their racism, or of preventing their infection in the first place, even if that means that some citizens will, because of their racism, come to believe that a fellow citizen is about to kill him when in fact he is not. Liberal states can use the criminal law in order to punish acts or omissions that cause or risk causing harms, but they cannot use it in order to punish acts or omissions that cause or risk causing the possession or retention of beliefs and desires, however illiberal the content of those beliefs and desires may be. Nor does it matter whether the punishing is done directly through rules of liability, or indirectly through forfeiture rules like the reasonable-belief rule.

CONCLUSION

Cases like Goetz* leave the liberal state in an awkward position. On the one hand, it does not want to acquit Goetz* because it does not want its citizens to believe that it condones the racism that caused him to believe that he was about to be killed. On the other hand, it can rely upon none of the theories examined above as a basis to deny him an acquittal. Goetz* cannot be punished for the killing alone, since he killed only because he believed he was about to be killed, and but for the reasonable-belief rule, the law permits him to kill under those circumstances. Nor can he be punished for forming the belief that he was about to be killed. He had no control over that. Nor can he be punished for being a racist, or for choosing to become or remain one. Liberal states do not punish people for who they are, nor do they punish them for choosing to become or remain who they are.

112. See Kang, *supra* note 105, at 1580, 1585 (describing “numerous variations on a strategy of debiasing public service announcements (d-PSAs)” meant to “counter [the] implicit [biasing] fire [of local news] with implicit [debiasing] fire”).

If an actor kills only because he believed that he was about to be killed, and if he believed that he was about to be killed only because he was a racist, we can and should condemn the racism that led to the belief. Citizens of liberal states should not be racists. Nonetheless, a liberal state has no basis upon which it can legitimately say that such an actor forfeits his claim of self-defense. Punishing an actor like Goetz* is not the liberal way to get to a liberal society. On the contrary, foregoing the punishment of such an actor is the price one pays for a liberal society in which the only legitimate basis upon which the state can punish a citizen is his choice to cause or risk causing harm when that choice is one the law does not permit him to make.