

# Racism, Unreasonable Belief, and Bernhard Goetz\*

*Stephen P. Garvey\**

## Abstract

How should the law 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 so believed if V had been white? Should D be exonerated on grounds of self-defense? The canonical case raising this question is *People v. Goetz*. Some commentators argue that norms of equal treatment and anti-discrimination require that D's claim of self-defense be rejected.

I argue that denying D's claim of self-defense would be at odds with the principle that criminal liability should only be imposed on an actor if he culpably chooses to cause or unjustifiably risk causing harm, not for possessing or choosing to possess racist or otherwise illiberal beliefs or desires. Moreover, insofar as this harm 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, norms of equal treatment and anti-discrimination to the contrary notwithstanding.

---

\* Professor of Law, Cornell Law School. Thanks to John Blume, Steven Clymer, 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.

## Table of Contents

Introduction .....	1
I. The Reasonable-Belief Rule .....	14
A. Forfeiture Rules .....	19
B. The Reasonable-Belief Rule as a Forfeiture Rule .....	24
II. The Character Theory .....	31
A. Is Racism Wrong? .....	32
1. Racism in the Heart .....	32
2. Racism in the Head .....	35
B. Punishing Being a Racist .....	40
III. The Belief-Choice and Character-Choice Theories .....	43
A. The Belief-Choice Theory .....	46
B. The Character-Choice Theory .....	49
C. Punishing the Choice to Be a Racist .....	50
1. Luck .....	51
2. Legality .....	52
3. Liberalism .....	56
Conclusion .....	57

## INTRODUCTION

Bernhard Goetz, a 37-year old white man with a “slight[] build,”<sup>1</sup> boarded a New York subway three days before Christmas in December 1984.<sup>2</sup> Four “noisy and boisterous”<sup>3</sup> young men,<sup>4</sup> all of them black,<sup>5</sup> were also on-board. Two of the young men approached Goetz.<sup>6</sup> “Give me five dollars,”<sup>7</sup> one of them said. When the request was repeated,<sup>8</sup> Goetz opened fire with a concealed .38 caliber pistol loaded with five rounds.<sup>9</sup> He fired four shots in rapid succession, wounding three of the

---

<sup>1</sup> GEORGE P. FLETCHER, *A CRIME OF SELF-DEFENSE: BERNHARD GOETZ AND THE LAW ON TRIAL* 1 (1988).

<sup>2</sup> 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 mainly on pre-trial statements Goetz made to the police. *People v. Goetz*, 497 N.E.2d 41, 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*, 131 Misc.2d 1 (Sup. Ct.), *aff’d*, 501 N.Y.2d 326 (App. Div.), *aff’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, 782 (App. Div.), *aff’d*, 532 N.E.2d 1273 (N.Y. 1988); Ronald Sullivan, *Goetz Is Given One-Year Term on Gun Charge*, N.Y. TIMES, Jan. 14, 1989, at B1.

Book-length accounts of the case include FLETCHER, *supra* note 1; MARK LESLY WITH 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).

<sup>3</sup> *See* FLETCHER, *supra* note 1, at 1.

<sup>4</sup> Two of the victims (James Ramseur and Barry Allen) were eighteen. The other two (Darell Cabey and Troy Canty) were nineteen. *See id.* at 2-3.

<sup>5</sup> *See id.* at 1.

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

<sup>7</sup> *Id.*

<sup>8</sup> *See id.* at 44.

<sup>9</sup> *See id.* at 43.

victims.<sup>10</sup> After pausing to survey the scene,<sup>11</sup> Goetz approached the fourth victim and said to him, “You seem to be all right, here’s another.” He then fired the final round.<sup>12</sup> The bullet severed the victim’s spinal cord, leaving him paralyzed.<sup>13</sup> Most of the other passengers fled when the shooting started, but two women, one of whom was black,<sup>14</sup> remained, “immobilized by fear.<sup>15</sup>” Goetz “sa[id] some soothing words”<sup>16</sup> to them, and jumped off the train. He later turned himself in.<sup>17</sup>

Charged with attempted second-degree murder,<sup>18</sup> Goetz claimed that he acted in self-defense, according to which an actor is, generally speaking, permitted to use deadly force if — but only if — he *reasonably* believed that the use of deadly force was necessary to avoid death or grievous bodily harm.<sup>19</sup> In other words, what matters to the law of self-

<sup>10</sup> *See id.*

<sup>11</sup> *See id.* 44. *But cf.* FLETCHER, *supra* note 1, at 171 (noting that “eight witnesses testified that they did not hear a pause”).

<sup>12</sup> *See id.*

<sup>13</sup> *See id.* The fourth victim was Darell Cabey. Cabey later filed a civil suit against Goetz and won a \$43 million judgment in 1996. A copy of the complaint can be found at <<http://www.lectlaw.com/files/cas91.htm>>.

<sup>14</sup> *See* FLETCHER, *supra* note 1, at 5.

<sup>15</sup> *Id.* at 2.

<sup>16</sup> *Id.*

<sup>17</sup> *See Goetz*, 497 N.E.2d at 44.

<sup>18</sup> *See id.* at 43.

<sup>19</sup> *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. LAFAVE, 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.”); 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.”). *But cf.* 2 PAUL H. ROBINSON, CRIMINAL LAW DEFENSES § 184(b)(1), at 399-400 (Supp. 2005-06) (“Whether any honest ‘belief’ or only a ‘reasonable belief’ will provide the defense will depend upon the specific statute or case law; there is much variation on this point.”).

defense is not what the defendant believed were the facts.<sup>20</sup> Nor do the facts themselves matter.<sup>21</sup> An actor who kills because he reasonably believes that he is about to be killed will still be acquitted on grounds of self-defense even if it later turns out that he was wrong. Again, what matters is what the actor reasonably believed. Call this the *reasonable-belief rule*.

Reactions to the case run from one extreme to the other.<sup>22</sup> At one end are those who would reject a claim of self-defense. Indeed, some in this camp don't see the case as one of self-defense at all.<sup>23</sup> On the

---

<sup>20</sup> An actor who kills because he honestly but unreasonably believed he was about to be killed may be entitled to a partial defense usually described as "imperfect self-defense." *See infra* p. 18.

<sup>21</sup> The facts may matter as to whether the defense is characterized as an excuse or as a justification. *See infra* p. 17.

<sup>22</sup> The jury acquitted on all the charges (including the attempted murder charges), except the firearms-possession charge. 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 reached this conclusion because they "incorporated Goetz's purpose of defending himself into their analysis of his intention [to kill]." *Id.* 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 shooting. Although this way of thinking about self-defense is generally at odds with contemporary thinking about the case, *see, e.g.,* FLETCHER, *supra*, at 186, it nonetheless has a long history rooted in the Catholic doctrine of double effect. *See, e.g.,* SUZANNE UNIACKE, PERMISSIBLE KILLING: THE SELF-DEFENSE JUSTIFICATION OF HOMICIDE 92 (1994) ("Homicide in self-defense has been characterized as unintended killing, and said to be justified under the conditions of the Principle of Double Effect."). The underlying philosophical question deals with how one should go about individuating the objects of intention. *See, e.g.,* MICHAEL MOORE, PLACING BLAME 469 (1997) (arguing in favor of a "very fine-grained" theory of individuation).

<sup>23</sup> *See infra* p. 32 (explaining why the Goetz case should not be seen as a hate crime). Then-U.S. Attorney Rudolph Giuliani refused to prosecute Goetz for "having deprived the four youths of their civil rights . . . [because there was] insufficient evidence that the shooting expressed a racial motive." FLETCHER, *supra* note 1, at 5.

contrary, they see it as a hate crime. Far from being acquitted, Goetz should have been convicted of attempted murder, and his sentence for that crime should have been enhanced for its hate.<sup>24</sup> Others don't see a hate crime, but neither do they think that Goetz believed his life was in danger. His claim to the contrary was a lie or rationalization. Still others think Goetz that did believe his life was in danger, but would reject his claim of self-defense all the same, believing it was unreasonable for him to have so believed. It might have been reasonable for him to have believed that he was in *some* danger of *some* harm, but not imminent death, nor even imminent serious bodily injury.<sup>25</sup>

At the other end are those who would accept a self-defense claim. Some in this camp join only because of an idiosyncrasy of New York self-defense law. They point out that New York law permits a person to use deadly force, not only to avoid being killed or seriously injured,<sup>26</sup> but also to avoid being robbed,<sup>27</sup> even if the robbery involves no threat of death or serious bodily injury.<sup>28</sup> Although they don't think Goetz reasonably believed that he was about to be killed or seriously injured, they do think he reasonably believed that he was about to be robbed, and

---

<sup>24</sup> At the time of the crime, New York law did not authorize sentence enhancements for hate crime. Current New York law does provide for such enhancements. See N.Y. PENAL CODE § 485.05 (McKinney 2006) (defining "hate crime"); *id.* at § 485.10 (providing for sentencing enhancement).

<sup>25</sup> Using deadly force in response to a non-deadly threat would provide the basis for a partial defense in some jurisdictions. See, e.g., 2 ROBINSON, *supra* note 19, § 131(c), at 77 ("Under the doctrine of 'imperfect self-defense,' . . . many jurisdictions impose liability for manslaughter where the defendant kills unnecessarily while resisting an unjustified attack.").

<sup>26</sup> N.Y. PENAL CODE § 35.15(2)(a).

<sup>27</sup> *Id.* at § 35.15(2)(b). See also 2 ROBINSON, *supra* note 19, § 131(d)(2), at 83 ("Some states expressly authorize the use of deadly defensive force in response to certain enumerated felonies.").

<sup>28</sup> N.Y. PENAL CODE § 160.00 ("Robbery is forcible stealing. A person forcibly steals property and commits robbery when, in the course of committing a larceny, he uses or threatens the immediate use of *physical force* upon another person[.]" whether or not that force amounts to deadly force) (emphasis added).

New York law, rightly or wrongly, allowed him to use deadly force to stop it. Others accept Goetz's claim of self-defense despite the idiosyncracies of New York self-defense law. They think not only that Goetz believed that he was about to be killed, they also think it was reasonable for him to have so believed. Consequently, he had every right to defend himself, even under more traditional formulations of the defense.<sup>29</sup>

The Goetz case is usually thought to raise three questions about the doctrine of self-defense. First, should an actor be entitled to claim self-defense *whenever* he believes that the use of deadly force is necessary to avoid death or grievous bodily harm — which for convenience sake will hereafter be referred to as the belief that *p* — whether or not that belief was a reasonable one for him to have held under the circumstances? Or should the defense be available only when an actor's belief that *p* is reasonable, as the reasonable-belief rule says? This question is largely academic, since most American jurisdictions adhere to the reasonable-belief rule.<sup>30</sup> Indeed, the rule is seldom questioned

---

<sup>29</sup> Some might argue that it was unreasonable for Goetz to have believed he was about to be killed based on the evidence available to him when he pulled the trigger, but that he was nonetheless in fact justified in using deadly force based on evidence of which he was unaware; to wit, that two of the youths were carrying screwdrivers used to break into the coin boxes of video machines, *see Goetz*, 497 N.E.2d at 43, and that some of the victims had in fact intended to rob him. *See id.* at 45 (describing this evidence). If so, then Goetz might be described as “unknowingly justified.” A lively debate exists over whether an actor who kills another person, who is in fact justified in killing the other person, but who is unaware of the facts establishing that justification, should be punished for murder, or only for attempted murder, or for nothing at all. *See, e.g.*, Anthony M. Dillof, *Unraveling Unknowing Justification*, 77 NOTRE DAME L. REV. 1547, 1554-56 (2002) (describing these three approaches).

<sup>30</sup> *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).

even among academic commentators.<sup>31</sup>

Second, what standard should be used to decide if an actor's belief that *p* is reasonable or unreasonable? Here the debate is usually framed in terms of what particular characteristics or features of the defendant or

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 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").

<sup>31</sup> For one exception, see Singer, *supra* note 30, 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"). Glanville Williams argued that an honest mistake should suffice as a defense to murder, but not involuntary manslaughter. See GLANVILLE WILLIAMS, *CRIMINAL LAW: THE GENERAL PART* § 71, at 201 (2d ed. 1961) (stating that the "idea that a mistake, to be a defence, must be reasonable . . . is . . . untenabl[e]" but going on to note that an actor who makes a "grossly unreasonable" mistake "may of course be convicted of manslaughter") [hereinafter WILLIAMS, *CRIMINAL LAW*]; GLANVILLE WILLIAMS, *TEXTBOOK OF CRIMINAL LAW* § 6.8, at 138 (2d ed. 1983) ("[I]f a person kills another in the convinced but mistaken and unreasonable belief that he himself is about to be killed by the other, he is theoretically guilty of murder, even though on the facts as he believed them to be he would not have been guilty of any crime. Is this not a harsh rule?");

his situation should be imputed to the reasonable person.<sup>32</sup> Those who favor a more “objective” test or standard urge relatively less “subjectification” of the reasonable person, while those who favor a more “subjective” test or standard urge relatively more. The second question is related to the first, inasmuch as a fully subjective rendering of the reasonable-belief rule would impute all the characteristics of the defendant and his situation to the reasonable person,<sup>33</sup> which would end up doing away with the rule altogether.<sup>34</sup> The reasonable person would

---

<sup>32</sup> 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?”).

<sup>33</sup> See, e.g., *State v. Leidholm*, 334 N.W.2d 811, 818 (N.D. 1983) (“[A] correct statement of the law of self-defense is one in which the court directs the jury to assume the physical and psychological properties peculiar to the accused.”).

<sup>34</sup> See, e.g., GEORGE 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.”); PAUL H. ROBINSON & MICHAEL T. CAHILL, LAW WITHOUT JUSTICE: WHY CRIMINAL LAW DOESN’T GIVE PEOPLE WHAT THEY DESERVE 50 (2006) (“[A] *complete* individualization of the objective standard . . . would produce a purely subjective standard.”).

One might argue that a fully subjective standard does not 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  $\phi$ ’s at the same time that he believes all things considered that he should not  $\phi$ . A debate exists as to whether this epistemic state is conceptually impossible or merely irrational. Compare Johnathan 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

have believed whatever the defendant in fact believed, because the reasonable person *is* the defendant.

Third, if the reasonable-belief rule should be applied with some degree of subjectification, but not complete subjectification, should the characteristics to be imputed to the reasonable person include the defendant's racism? Or to put the question more provocatively: Is the reasonable person a racist? For some, the answer is obvious: Of course not. But the harder question is: Why not? How do we know which of the accused's characteristics to impute to the reasonable person and which to exclude? If the criminal law can provide an answer to that question,<sup>35</sup> it has yet to do so, preferring instead to let each jury provide its own answer one case at a time.<sup>36</sup> Juries are told that the defendant's belief that *p* must be reasonable, but they are seldom told much more

---

irrational, but incoherent.”); David Owens, *Epistemic Akrasia* 85 *THE MONIST* 381, 395 (2002) (“[E]pistemic akrasia is not possible.”), 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 Mele, *Incontinent Believing*, 36 *PHIL. Q.* 212, 217 (1986) (arguing that “full-blown incontinent believing” is “possible”).

<sup>35</sup> Larry Alexander has argued that any answer to the question 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. He elaborates:

[A]ny RPAS [reasonable person in the actor's situation] will be a construct that includes some beliefs of the actual defendant together with the beliefs that the constructor inserts. Which beliefs are inserted other than the ones the defendant actually had will determine whether or not the RPAS would act as the defendant acted. But there is no standard that tells us which of the beliefs of the actual defendant should be left intact and which should be replaced by other (correct) beliefs.

*Id.*

<sup>36</sup> See, e.g., ROBINSON & CAHILL, *supra* note 34, at 49-51 (noting that “criminal-law theorists have not yet been able to articulate a comprehensive principle that defines what should and should not 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”).

than that.

My argument begins from the premise that an actor should only be punished if he chooses to do that which the law does not permit him to do. An actor who kills only because he unreasonably believes that he is about to be killed does *not* choose to do that which the law does not permit him to do. On the contrary, what *he* chooses to do is to exercise deadly force because it was necessary to do so, but the use of deadly force under those circumstances is something that, save for the reasonable belief rule, the law would permit him to do. The only reason the actor kills is because he believes, albeit unreasonably, that he is about to be killed, and like us all, he is hostage to his beliefs at the moment he acts. We can only guide our actions based on the beliefs we hold at the time we act, and the belief that *p* — that the use of deadly force is necessary to avoid death or serious bodily injury — is one that only a saint or a fool would fail to heed. Consequently, I would argue that self-defense should be available to any actor who kills because he believed that doing so was necessary to avoid being killed or seriously injured, whether that belief is reasonable or not.<sup>37</sup>

Nonetheless, I make no argument here for abolishing the reasonable-belief rule. Instead, I argue in Part I that the rule should be understood as a forfeiture rule. On this view, when an actor kills because and only because he believes that *p*, he forfeits the defense of self-defense if and because his belief that *p* is said to be “unreasonable.” But

---

<sup>37</sup> Would such an approach to the law of self-defense result in less protection for innocent victims of erroneously-deployed self-defensive force? It is hard to see how. Would an actor’s awareness of the reasonable-belief rule stop him from using deadly force if and when the actor came truly to believe that he was about to be killed? If not, then 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. *Cf.* 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.”). Moreover, can an actor believe that a belief he holds is unreasonable and rationally continue to hold it? *See supra* note 34. If not, then so long as an actor is rational the reasonable-belief rule will have no effect on his actions: an actor who believes he is about to be killed will also believe that that belief is reasonable.

describing an actor's belief as "unreasonable" is the law's way of saying that the actor violated an obligation but-for which he would not have believed that *p*. Consequently, if Geotz attempted to kill because and only because he believed that he was about to be killed, and if the law nonetheless convicts him attempted 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 violated some other obligation. His belief that *p* is therefore "unreasonable" in the eyes of the law.

The Goetz case is controversial. It elicits strong reactions and so might be considered a poor vehicle for examining the reasonable-belief rule. Why not choose a case less apt to touch a nerve? I take the point, but nonetheless focus on the case, or more precisely on a variation of the case, in part because it is so controversial, and in part because it has entered the criminal-law canon.<sup>38</sup> Few American lawyers will have left law school without at some point having encountered it. In this connection, one group of authors notes that the "case became a cause célèbre."<sup>39</sup> Another author called it a "cultural monument."<sup>40</sup> In 2003,

---

<sup>38</sup> The case is reproduced in several criminal-law casebooks. See RICHARD J. BONNIE ET AL., CRIMINAL LAW 419 (2d ed. 2004); RONALD N. BOYCE, DONALD A. DRIPPS & ROLLIN M. PERKINS, CRIMINAL LAW AND PROCEDURE: CASES AND MATERIALS 966 (9th ed. 2004); JOSEPH G. COOK & PAUL MARCUS, CRIMINAL LAW 691 (4th ed. 1999); JOSHUA DRESSLER, CASES AND MATERIALS ON CRIMINAL LAW 495 (3d ed. 2003); MARKUS D. DUBBER & MARK G. KELMAN, AMERICAN CRIMINAL LAW: CASES, STATUTES, AND COMMENTS 542 (2005); SANFORD H. KADISH & STEPHEN J. SHULHOFER, CRIMINAL LAW AND ITS PROCESSES: CASES AND MATERIALS 801 (6th ed. 1995); JOHN KAPLAN, ROBERT WEISBERG & GUYORA BINDER, CRIMINAL LAW: CASES AND MATERIALS 521 (5th ed. 2004); WAYNE R. LAFAVE, MODERN CRIMINAL LAW: CASES, COMMENTS AND QUESTIONS 510 (4th ed. 2006); CYNTHIA LEE & ANGELA HARRIS, CRIMINAL LAW: CASES AND MATERIALS 425 (2005); ANDRE A. MOENSSENS, CRIMINAL LAW: CASES AND COMMENTS 518 (7th ed. 2003); PAUL ROBINSON, CRIMINAL LAW: CASE STUDIES AND CONTROVERSIES 559 (2005).

<sup>39</sup> KADISH ET AL., *supra* note 38, at 806.

<sup>40</sup> FLETCHER, *supra* note 1, at xi. *But see* Franklin E. Zimring, *Hardly the Trial of the Century*, 87 MICH. L. REV. 1307, 1307 (1989) (book review) ("[W]hat is there about [the Goetz case] that justifies its landmark status in public discussions of crime and

*New York Magazine* included Goetz among the “100 People Who Changed New York.”<sup>41</sup> Moreover, my goal is not really to understand the Goetz case at all, let alone to defend the real Bernhard Goetz. 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 *p* is reasonable.

In order to accomplish that goal without needless distraction related to the details of the real Goetz case, I ask you to imagine from here on out that Goetz fired only one shot from his .38, not all five; that he hit only one of the young men, not all four; that the other men fled unharmed; and that Goetz fired because and only because he believed he was about to be killed.<sup>42</sup> In order to avoid any unnecessary doctrinal complications arising from the special mens rea requirements associated with the law of attempts,<sup>43</sup> I also ask you to imagine that Goetz’s first and only shot killed its intended victim, rather than simply wounding him. My references to Goetz from this point forward should be understood as references to this stripped-down version of the facts, and not to the messier facts of the actual case.<sup>44</sup> Moreover, in an effort to forestall

---

criminal justice? Perhaps there is less than we might suppose.”).

<sup>41</sup> *100 People Who Changed New York*, N.Y. MAG., Apr. 7-14, 2003, at 98, 99.

<sup>42</sup> For many people, Goetz transformed himself from a potentially sympathetic victim who might have been acting in self-defense, and into a decidedly unsympathetic vigilante acting in retaliation, when, at least according to his confession, he paused, stood over his final victim, and fired the shot severing his spinal cord. See FLETCHER, *supra* note 1, 170. *But see id.* at 171 (noting that the testimony of other witnesses suggested no such pause); Singer, *supra* note 30, at 516 (“[E]ven Goetz’s fifth shot could be found to have emanated from a swirl of anxiety and loss of control which continued after the last shot.”).

<sup>43</sup> See, e.g., DRESSLER, *supra* note 19, § 27.05, at 417-22 (discussing the mens rea of attempts).

<sup>44</sup> Because I focus on this much-sanitized version of the case, some of my colleagues have accused me of a bait-and-switch: I lure the reader into thinking I will be discussing the Goetz case, when in fact I am discussing a wholly different case. Perhaps. But even if one agrees that Goetz should not have been acquitted on grounds of self-defense with respect to the attempted murder of the fourth victim, what about the first?

any confusion, I will refer to the Goetz of my simplified case as Goetz\*.

For those who think it was unreasonable for Goetz\* to believe that  $p$ , the unreasonableness of his belief usually has something to do with the idea that he was a racist. Accordingly, I assume for present purposes that Goetz\* was indeed a racist,<sup>45</sup> and that he would not have believed that  $p$  had he not been a racist. In other words, had the victims been white, Goetz\* would not have believed he was about to be killed.<sup>46</sup> With those

I fail to see any material difference between Goetz's conduct with respect to the first victim, or perhaps even the first three, and the conduct of the actor in my sanitized version of the case. But I would guess that those who believe themselves victims of a bait-and-switch are disinclined not only to acquit Goetz of the fourth attempt, but of the first three as well.

<sup>45</sup> Having made this assumption, I would nonetheless note that according to an article written in *New York Magazine* soon after the first grand jury declined to indict 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. At Goetz's eventual trial, the trial judge refused to permit the prosecution to admit Goetz's racist comment into evidence. See FLETCHER, *supra* note 1, at 204. According to George Fletcher, who observed the proceedings against Goetz first-hand, "[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." *Id.* at 205

<sup>46</sup> 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 AND SOC. PSYCHOL.* 13 (2002); Charles M. Judd *et al.*, *Automatic Stereotype v. Automatic Prejudice: Sorting Out the Possibilities in the Payne (2001) Weapons Paradigm*, 40 *J. EXPERIMENTAL SOC. PSYCH.* 75 (2004); B. Keith Payne, *Prejudice and Perception: The Role of Automatic and*

assumptions in hand, at least three different theories can be offered to explain why Goetz\*'s belief that *p* was unreasonable, each of which sets forth a duty that Goetz\* breached and upon which the forfeiture of his self-defense claim is based.

According to the *character theory* (discussed in Part II), 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 traces the unreasonableness of Goetz\*'s belief that *p* to his racist character. Goetz\* forfeits his claim of self-defense because he violated the duty not to be a racist.

The second and third theories (discussed in Part III) trace the unreasonableness of Goetz\*'s belief that *p* to a choice that he made. According to the *belief-choice theory*, Goetz\*'s belief that *p* was unreasonable because he chose to believe that *p* when he could have chosen otherwise, and he should have chosen otherwise inasmuch as his choice to believe that *p* was based on racism. Goetz\* forfeits his claim of self-defense because he violated the duty not to choose to believe based on racism. 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 when he could and should have chosen otherwise, and his racism in turn caused him to form the belief that *p*. Goetz\* forfeits his claim of self-defense because he violated the duty not to choose to become or remain a racist.

I argue that a liberal state can embrace none of these theories. In one way or another each is inconsistent with the principle that an actor should only be punished if he chooses through his acts or omissions 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

---

*Controlled Processes in Misperceiving a Weapon*, 81 J. PERSONALITY AND SOC. PSYCHOL. 181 (2001); 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. PSYCH. 384 (2002).

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. If acquitting Goetz\* on grounds of self-defense is for some reason believed to be the wrong result, it is the result to which a liberal state should nonetheless be committed, unless it wishes to abandon the above-mentioned principle, or until some other theory consistent with that principle is forthcoming.<sup>47</sup>

#### I. THE REASONABLE-BELIEF RULE

As a general proposition an actor who uses deadly force against another can claim self-defense if he reasonably believed that the use of such force was necessary to prevent his death or serious bodily injury. In some jurisdictions the actor must also have reasonably believed that his assailant's use of such force was imminent,<sup>48</sup> or immediately neces

---

<sup>47</sup> I want to emphasize that I make no claim that the theories examined here are exhaustive. All I hope to have accomplished is to place the argumentative burden of proof on those who believe Goetz\* can be punished consistent with the principle mentioned in the text.

<sup>48</sup> One can imagine cases in which an actor reasonably believes he is facing imminent death or serious bodily injury but nonetheless unreasonably believes 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 DRESSLER, *supra* note 19, § 18.02[C], at 238.

For arguments in favor of eliminating imminence 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 we may properly defend against from mere inchoate and potential threats”).

No jurisdiction 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

sary.<sup>49</sup> If the actor's beliefs regarding the elements of the defense were reasonable,<sup>50</sup> he is entitled to what is sometimes called "perfect" self-

---

statute defining self-defense says, for example, that an actor must reasonably believe that the probability of death or serious bodily injury is 75% 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.

<sup>49</sup> MODEL PENAL CODE § 3.04(1).

<sup>50</sup> The Model Penal Code's self-defense provision does not speak in terms of "reasonable" or "unreasonable" beliefs, *see* ROBINSON, *supra* note 19, § 4.4, at 263-64 (describing differences between common law and MPC approaches to mistakes), although the Code's general definitions section does say that the term "reasonable belief" designates a belief that the actor is not reckless or negligent in holding." Instead, the Code's self-defense provision says that an actor is permitted to use deadly force if he believes — reasonably or not — 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). 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 that 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.* at § 3.09(2). This provision is usually understood to mean that an actor who recklessly believes that the use of deadly force is 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 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

defense, which is a complete or full defense, even if it turns out later that he was wrong to believe that he was about to be attacked. In other words, an actor is permitted to kill if and when he reasonably believes he is about to be killed, or he reasonably believes that the only way he can avoid being killed is to kill.<sup>51</sup>

---

believes that  $p$  but at the same time suspects that not- $p$ , where not- $p$  is true. See, e.g., Douglas N. Husak & Craig A. Callender, *Wilful Ignorance, Knowledge, and the "Equal Cul- pability" 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- $p$ , where not- $p$  is true. See, e.g., Larry Alexander, *Lesser Evils: A Closer Look at the Paradigmatic Justification*, 24 LAW & PHIL. 611, 624-25. This approach treats an actor who recklessly believes that  $p$  as someone suffering from epistemic akrasia. See *supra* note 34. For present purposes, I will continue to speak in the more common common-law terminology of reasonable and unreasonable belief.

<sup>51</sup> Mark Kelman and Jody Armour 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 Armour, *Race Ipsa Loquitur: Of Reasonable Racists, Intelligent Bayesians, 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 the probability of him being killed (unless he killed first) was 75%. According to the Kelman-Armour thesis, Goetz should not, despite his belief, be entitled to self-defense if a jury decides that the false-negative error costs associated with requiring Goetz 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. Because Goetz "selected [his victims] on the basis of their race," Kelman, *supra*, at 815, those costs also include the stigmatization of 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%, 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 permits an actor to use deadly force when he reasonably believes that the use of

Under New York law an actor “may not use deadly force upon another person . . . unless [among other things]: [t]he actor reasonably believe[d] that such other person [was] using or about to use deadly physical force.”<sup>52</sup> Conversely, an actor is permitted to use deadly force upon another person if he reasonably believed that such other person [was] using or about to use deadly physical force against him. Moreover, New York law also permits an actor to use deadly force if he “reasonably believe[d] that such other person is committing or attempting to commit a kidnapping, forcible rape, forcible sexual assault or robbery,”<sup>53</sup> even if the other person, though committing or attempting to commit one of these offenses, is not using or about to use deadly physical force. Nonetheless, for present purposes I will assume that Goetz\* was only permitted to use deadly force if he reasonably believed that his victim was using or about to use deadly force against him. Again, for convenience sake, I will refer to this belief as the belief that *p*.

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. At the very least, the law permits

---

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 of him being killed or seriously injured (unless he kills first) is  $\Phi$ , 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 makes reference only to the reasonableness of the defendant’s factual perceptions, we *in fact* also expect the jury to assess 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”) (emphasis added). 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 when he reasonably believes that doing so is necessary. He cannot use deadly force against non-deadly force, even when the use of deadly force is necessary to avoid non-deadly injury.

<sup>52</sup> N.Y. PENAL CODE § 35.15(2)(a).

<sup>53</sup> *Id.* § 35.15(2)(b).

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 of considerable controversy. In one camp are those who argue that such an actor, like the actor who reasonably and correctly believes that *p*, is justified in using deadly force.<sup>54</sup> For those in this camp, self-defense is always a justification. In the opposing camp are those who argue that such an actor's use of deadly force is excused, but not justified.<sup>55</sup> For those in this camp, self-defense is sometimes a justification, and sometimes an excuse.<sup>56</sup> It all depends on whether the actor's reasonable belief turns out to be true (justification) or not (excuse).

Likewise, everyone also agrees that an actor who unreasonably believes that *p* is neither justified nor excused. In some jurisdictions, in-

---

<sup>54</sup> See, e.g., 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 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 (1994); 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 (1984); Kent Greenawalt, *The Perplexing Borders of Justification and Excuse*, 84 COLUM. L. REV. 1897, 1903 (1984); Hamish Stewart, *The Role of Reasonableness in Self-Defense*, 16 CAN. J.L. & JURIS. 317, 336 (2003).

<sup>55</sup> See, e.g., FLETCHER, *supra* note 34, § 10.1, 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 45, 47 (A.P. Simester & A.T.H. Smith eds., 1996); Paul Robinson, *Criminal Law Defenses: A Systematic Analysis*, 82 COLUM. L. REV. 199, 239-40 (1982).

<sup>56</sup> 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 (1996). According to another, the entire excuse-justification debate is misguided because 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 belief are unreasonable." *Id.* at 838 n.25.

cluding 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.”<sup>57</sup> As such, it is only a partial defense, mitigating to manslaughter what would otherwise be murder.

#### 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 to fail to 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 defense.<sup>58</sup> The general

---

<sup>57</sup> 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 Robinson & John M. Darley, *Testing Competing Theories of Justification*, 76 N.C. L. REV. 1095, 1128 (1998).

<sup>58</sup> 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 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 choice to encounter a threat, but instead on negligently encountering 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). 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 usually an illegitimate basis upon which to premise retributive punishment, and because forfeiture rules impose retributive punishment for the act or omission constituting the basis for the

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 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 a defense to which he would otherwise have been entitled.

The so-called “aggressor rule” 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,”<sup>59</sup> forfeits any claim of self-defense to which he might otherwise have been permitted, provided that he provoked the use of such force against himself “with the purpose of causing death or serious bodily injury.”<sup>60</sup> 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.<sup>61</sup> For example, the Model Penal Code provides that an actor

---

forfeiture.

<sup>59</sup> MODEL PENAL CODE § 3.04(2)(b)(i).

<sup>60</sup> *Id.* 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 suggests 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 the 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 more controversial) formulations of the aggressor rule.

<sup>61</sup> Forfeiture rules are seldom attached to insanity-defense provisions. See, e.g., Paul 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

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 subjected to duress.”<sup>62</sup> 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.<sup>63</sup> Under these types of rules an actor forfeits a defense to which he would otherwise be 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.<sup>64</sup>

---

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 or volitional incapacity associated with traditional tests of insanity. For example, suppose that he chooses not to take medicine he knows he needs 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 (1989-90) (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”).

<sup>62</sup> MODEL PENAL CODE § 2.09(2). 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.

<sup>63</sup> *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.

<sup>64</sup> 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 one is taking. Characterizing the voluntary-intoxication rule as an evidentiary rule would mean that the state is still required to prove that the defendant realized 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

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, and someone gets killed, he is not guilty 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 is creating a 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.<sup>65</sup> He chose to drink, and unlucky for him, he happened to kill

---

rule would mean that the state is not required to prove the requisite awareness. Instead, it would mean that the state believes that the crime of 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. Egelhof*, 518 U.S. 37, 41 (1996) (plurality opinion) (characterizing the Montana rule at issue as a rule of 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 the defendant was insane, but it cannot be introduced to show that he lacked any mental state associated with the crime charged. *See* DRESSLER, *supra* note 19 § 26.02[B][3]-[4], at 396-98 (discussing “no-defense” approach to the “*mens rea* form of diminished capacity”). 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. A voluntarily-intoxicated actor who lacks *mens rea* because he is intoxicated is at least responsible for becoming intoxicated, whereas a mentally-ill actor who lacks *mens rea* because he is mentally ill is ordinarily responsible neither for his becoming mentally ill nor for the behavioral manifestations of his illness.

<sup>65</sup> *See, e.g.*, MODEL PENAL CODE § 2.08. For three slightly different interpretations of § 2.08, *see* Westen, *supra* note 64, 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).

Opponents of the voluntary-intoxication rule have proposed a separate crime of dangerous intoxication for which “conviction . . . should usually result in purely

someone while intoxicated, even though he never realized 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.

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 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 robbery. 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 punish actors for crimes they did not commit, and as such, forfeiture rules necessarily result in disproportionate punishment, though *how* disproportionate will of course depend on the facts.

Nonetheless, my goal here is not to criticize forfeiture rules because they punish people disproportionately.<sup>66</sup> Instead, my goal is to argue that the reasonable-belief rule of self-defense functions as a forfeiture rule, and that a liberal state cannot legitimately apply this rule in cases like that of Goetz\* without sacrificing some of its basic liberal commitments. I turn now to the first part of this task.

---

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, CRIMINAL LAW, *supra* note 31, § 183, at 573-74.

<sup>66</sup> See Robinson, *supra* note 61, at 28-29.

B. The Reasonable-Belief Rule as a Forfeiture Rule

Although the reasonable-belief rule is not usually portrayed as a forfeiture rule, that is what it is. The reasonable-belief rule is like the voluntary-intoxication rules inasmuch as both involve an impairment of the actor's mind. One difference between them is that the intoxicated actor's impairment takes the form of ignorance, whereas the unreasonably-believing actor's impairment takes the form of a mistake. The intoxicated actor fails to form a belief that he should have formed, and but-for his intoxication would have formed. He fails to see a risk that he should have seen. In contrast, the unreasonably-believing actor forms a belief that  $p$  when he should not have formed that belief, and but-for some prior breach of duty would not have formed it. He sees a risk he should not have seen.

The voluntary-intoxication rule says that an actor forfeits a failure-of-proof defense if and because he chose to get drunk. The reasonable-belief rule says that an actor forfeits a claim of self-defense if and because he unreasonably believes that  $p$ .<sup>67</sup> But what does it mean to say that an actor 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 encounter on the subway, 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

---

<sup>67</sup> 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 begs the question. It simply presupposes that the reasonable-belief rule is somehow intrinsic to the defense itself rather than a forfeiture rule extrinsic to it.

what is the case. That belief in turn caused, or may have caused, Goetz\* to think about what he ought to do. Either way, 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\* decided or 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 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*. 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."<sup>68</sup> Instead, I assume that the problem was with his evidence. The problem, one might say, was not with Goetz\*'s cognitive hardware, but with 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 partic-

---

<sup>68</sup> Goetz\* did not believe that *p* just because he believed that his putative assailants were black: *that* would be paranoid. Cf. AM. PSYCHIATRIC ASS'N, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 634 (4th ed. 1994) ("Individuals with . . . [paranoid personality disorder] assume that other people will exploit, harm, or deceive them, even if no evidence exists to support this expectation.").

ularly 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: males are more prone to violence than females; and the young are more prone to violence than the old. These beliefs are of course generalizations.<sup>69</sup> 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 otherwise. 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: that blacks are more prone to violence than non-blacks, 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.<sup>70</sup> It was, so to speak, in his head.

Commentary on the Goetz\* case tends to assume that Goetz\*'s racism was cognitive.<sup>71</sup> Goetz\* was a racist because he *believed* that  $q$ . But there is another possibility. Goetz\*'s racism may not have been in

---

<sup>69</sup> In fact all 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.

<sup>70</sup> See, e.g., Kwame Anthony Appiah, *Racisms*, in *ANATOMY OF RACISM* 3, 5 (David Theo Goldberg ed., 1990) (“[E]xtrinsic racists make moral distinctions between members of different races because they believe that racial essence entails certain morally relevant qualities.”).

<sup>71</sup> See, e.g., Armour, *supra* note 51, at 782 (racism consists in the actor’s belief that “blacks are more prone than whites to be criminals”); Kelman, *supra* note 51, at 812 (racism consists in the actor’s beliefs “about the criminal predilections of black teenagers”).

his head, but in his heart. In other words, his racism was conative,<sup>72</sup> not cognitive.<sup>73</sup> 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,<sup>74</sup> 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.<sup>75</sup> Call this desire, or relative lack of desire, the desire that *q*.<sup>76</sup>

---

<sup>72</sup> 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 other accounts 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*, PHIL. & SOC. CRITICISM 1 (1999); J.L.A. Garcia, *Racism and Racial Discourse*, 32 PHIL. FORUM 125 (2001). For criticism of Garcia's conative conception of racism, see, for example, 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").

<sup>73</sup> 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).

<sup>74</sup> See, e.g., Garcia, *Heart of Racism*, *supra* note 72, 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.").

<sup>75</sup> 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 way about people assigned to certain racial groups, where this disregard is based on racial classification.").

A number of criminal-law scholars have argued that indifference, whether race-based or otherwise, should play a more important role in criminal-law theory and doctrine than it now does. For example, they have argued that an actor who creates a risk of causing a prohibited harm, but who did so unwittingly, can still fairly be subject to retributive punishment if his unawareness 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 258 (2003) ("[T]he indifference account places its focus on the *attitude* displayed by any particular action."); SAMUEL H.

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 non-blacks without harboring any malice toward them; or he might harbor malice toward them without possessing any belief or set of beliefs that might rationalize or make sense of such a sentiment.<sup>77</sup>

---

PILLSBURY, *JUDGING EVIL* 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?*, 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. LEG. 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. Simister ed., 2001) (arguing that 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, for example, Larry Alexander, *Insufficient Concern: A Unified Conception of Criminal Culpability*, 88 CAL. L. REV. 931, 938 (2000); Stephen P. Garvey, *What’s Wrong With Voluntary Manslaughter?*, 85 TEX. L. REV. 333, 357-63 (2006).

<sup>76</sup> 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.

<sup>77</sup> See, e.g., BLUM, *supra* note 73, at 10 (“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.”).

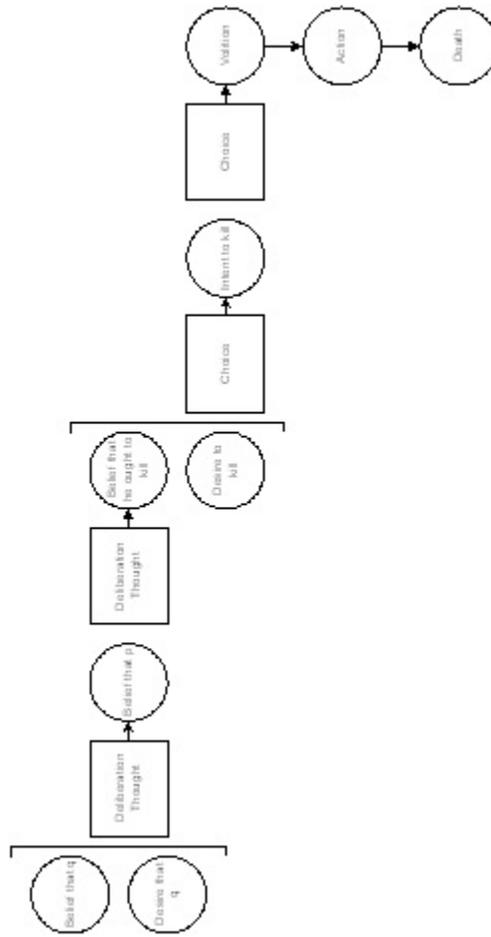
The important point for now is that the beliefs an actor forms at any moment in time depends not only on the background beliefs he possesses at that moment, but also on the background desires he possesses. When some desire other than the desire to discover the truth and avoid error influences what an actor believes, we might describe the actor as self-deceived, or say that his belief is the product of wishful thinking.<sup>78</sup> 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 common-sense psychology as follows:<sup>79</sup>

---

<sup>78</sup> See, e.g., ALFRED MELE, SELF-DECEPTION UNMASKED 50-51 (2001) (offering an analysis of self-deception according to which among other things an actor is self-deceived if his desires cause him to form a false belief); Robert Audi, *Self-Deception, Rationalization and the Ethics of Belief*, in ROBERT AUDI, MORAL KNOWLEDGE AND ETHICAL CHARACTER 132 (1997) (offering an analysis of self-deception according to which among other things an actor is self-deceived if, though believing that which is true, desire pushes that belief into unconsciousness, such that the actor sincerely avows that which is false); 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 largely because they want to believe" as they do).

<sup>79</sup> 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, and that Goetz\* should therefore forfeit his claim of self-defense, 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, it must be the case that liberal citizens ought not to possess the racist beliefs or desires we are assuming

that Goetz\* possessed. Second, it must be the case that a liberal state can 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 forfeiture. 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 is straightforwardly inconsistent with citizenship in a liberal state. Racism rooted in an actor's beliefs 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 the 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 is already a criminal offense, what would otherwise be a run-of-the-mill crime turns into a hate crime.<sup>80</sup> Although the subject is one of

---

<sup>80</sup> 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 body-acts — are fair targets of criminalization insofar as an actor has control over whether 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

considerable controversy, a hate crime is probably best analyzed as an ordinary crime committed with a specific goal or purpose in mind.<sup>81</sup> An otherwise criminal act 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

---

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 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).

<sup>81</sup> See, e.g., Kent Greenawalt, *Reflections on Justifications for Defining Crimes by the Category of Victim*, 1992/1993 ANN. SURV. AM. L. 617, 620-25; Jeffrie G. Murphy, *Bias Crimes: What Do Haters Deserve?*, 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 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).

*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 wasn't 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.<sup>82</sup> 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. The reasonable person is a normative ideal, not just a mirror image of our baser selves.<sup>83</sup> It holds us to a moral norm, not just a statistical one. The simple fact of the matter is that a liberal citizen does not harbor race-based 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.

---

<sup>82</sup> See, e.g., Armour, *supra* note 51, 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.”).

<sup>83</sup> See, e.g., LEE, *supra* note 32, at 235 (“[N]ormative reasonableness is a conception of reasonableness that focuses on the beliefs and actions society *ought* to recognize as reasonable. A positive (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.”).

## 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  $q$  — that blacks are more prone to violence than non-blacks — turns out to be more controversial. Racism in the heart is verboten 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, or was at least permitted to believe, that  $q$  because most people believe that  $q$ , assuming that they do.<sup>84</sup> What most people believe is once again 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  $p$ , or was at least permitted to believe that  $p$ , boils down to the claim that the proposition that  $q$  is true, and one ought to believe the truth, not to mention being permitted to believe it. Moreover, one can hardly be called a racist for believing the truth. One can imagine two responses to this claim.

First, 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.<sup>85</sup> Instead, it is a false or misleading generalization, or in other words, a

---

<sup>84</sup> According to a 1985 New York Times/WCBS New Poll, “[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 white was 55 percent and 71 percent, respectively.” Howard Kurtz, *Goetz Sentenced to 6 Months for Subway Shooting*, WASH. POST, Oct. 20, 1987, at A8.

<sup>85</sup> Compare Armour, *supra* note 51, 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*, THE NEW REPUBLIC, Sept. 13 & 20, 1999, at 30, 32 (“Statistics abundantly confirm that African Americans — and particularly young black men — commit 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.”).

stereotype.<sup>86</sup> Or, even if the proposition 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.<sup>87</sup> In either case, if a proposition itself 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 possess the offending belief. Thus, Goetz\* should not have believed that  $q$ .

Second, although the proposition that  $q$  is true, and although it cannot, all else being equal, be wrong to believe that which is true, 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.<sup>88</sup> 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 non-black motorists. Call this proposition  $q^*$ . If  $q^*$  is true, then a police officer who relies on a driver's race when deciding who 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 none-

<sup>86</sup> 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.”).

<sup>87</sup> 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 51, at 791 (“[T]he typical person tends to perceive race as the *overriding* factor when the supposed assailant is black.”).

<sup>88</sup> 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.

theless have compelling reasons to prohibit police officers from relying on race when deciding who and who not to stop. Effective law enforcement is one goal worth pursuing, but not the only one. Profiles that include race threaten to increase racial stigmatization and social isolation of the group stigmatized.<sup>89</sup> 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 of the costs and benefits of viable alternatives.<sup>90</sup>

---

<sup>89</sup> Compare SCHAUER, *supra* note 87, 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.”); Kennedy, *supra* note 85, 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.”).

<sup>90</sup> Compare SCHAUER, *supra* note 87, 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 Gross & Debra Livingston, *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.”); Kennedy, *supra*, at 34 (“[O]ur 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.”); Peter Schuck, *A Case for Profiling*, AM. LAW., 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, for example, Samuel R. Gross & Katherine Y. Barnes, *Road Work: Racial Profiling and Drug Interdiction on*

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  $q$ ?<sup>91</sup> If  $q$  is 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. Again, the law might nonetheless have compelling reasons to prohibit an actor's reliance on race. Again, official recognition of  $p$  threatens to increase racial stigmatization and social isolation, and again, how one comes down on the question will depend on the size of these effects, and on the moral weight one assigns them.<sup>92</sup>

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

---

*the Highway*, 101 MICH. L. REV. 651, 744 (2002) (summarizing conclusions based on analysis under the Fourth Amendment and the Equal Protection Clause).

<sup>91</sup> This assumption is rejected in Part II.

<sup>92</sup> See, e.g., RANDALL KENNEDY, RACE, CRIME AND 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 51, at 795 ("Hastier use of force against blacks forced blacks who do not want to be mistaken for assailants to avoid ostensibly public places . . . and core community activities."); Kelman, *supra* note 51, 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 more about them when who they truly are as individuals is wholly misassessed.").

jury's search for the truth might be another.<sup>93</sup> 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. White defendants who 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 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, if Goetz\* did believe that  $p$ , then the belief upon which he "acted" was the belief that his life was in danger, and the desire upon which he "acted" was the desire to save his life. He did not "act upon" his racist belief or desire.<sup>94</sup> Those mental states entered the

---

<sup>93</sup> See Armour, *supra* note 51, at 49 ("[P]ermitting a defendant who shoots a 'suspicious' Black person to focus on race at trial, even for the ostensibly neutral purpose of supporting the rationality of his factual judgments, impairs the capacity of jurors rationally and fairly to strike the same balance."). For ideas about how the law might counteract the effects of prejudice on jury decision-making, see, for example, LEE, *supra* note 32, at 252-53 (proposing that judges give "race-switching" instructions in appropriate cases); 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.")

<sup>94</sup> When we say 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, e.g., David A. Straus, *Discriminatory Intent and the Taming of Brown*, 56 U. CHI. L. REV. 935, 956-59 (1989) (proposing this definition of "discriminatory intent"); 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

picture, not at the point of action, but earlier, at the point of belief formation. If one nonetheless insists that Goetz\* did act on the belief that  $q$  at the moment he pulled the trigger, then it must also be the case 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 we act on all our beliefs whenever we act on any of them. Thus, if the law refuses to credit Goetz\*'s claim of self-defense, it does so in the end because 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.

Can a liberal state deny an actor a defense to which he would otherwise have been entitled because he was a racist? Asking this question is equivalent to asking whether a liberal state could punish an actor for being a racist. If a liberal state cannot punish a person for being a racist, then neither should being a racist be the basis upon which an actor is denied a defense to which he would otherwise be entitled. If a state wants to condition access to a defense through the operation of a forfeiture rule, that rule should be one upon which the state could independently predicate criminal liability.

#### B. Punishing Being a Racist

Before the law can legitimately punish a person for something, the person must first be responsible for it. Was Goetz\* responsible for the

---

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 anti-discrimination law, including a “causal account.”). The actor may or may not be aware that he possesses the stereotype or that his unobjectionable belief is based on, or caused by, 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 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 a defense to criminal liability.

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 one theory of responsibility, known as the character theory, he was.

The character theory of responsibility 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 *we are* our characters.<sup>95</sup> In other words, we are responsible for who we are 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, condemn 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 the state can condemn him for being a racist through the suffering and hardship of *punishment*. A liberal state need not treat the racists in its midst as it does those who accord their fellow citizens the equal concern to which they are due. It is free to censure 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.<sup>96</sup> Nor should the content of an actor's

---

<sup>95</sup> See, e.g., MOORE, *supra* note 22, at 571 (“[W]e are responsible for our characters because we are, in part, constituted by our characters.”).

<sup>96</sup> 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*:

character form the basis upon which the state denies him a defense to which he would otherwise have been entitled.

The analogy here would be to so-called status offenses. Our law does not in fact punish actors just for being who they are, nor should it.<sup>97</sup> 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 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 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, ought nonetheless to receive the state's mercy, or

---

*Activity and Passivity in Mental Life*, 115 ETHICS 236, 270-71 (2005) (arguing that we are responsible for our beliefs but noting that “one 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 54, 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 require 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.”).

<sup>97</sup> 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 Eight Amendment because it “inflicts a cruel and unusual punishment”).

at least be a candidate for its mercy.<sup>98</sup> For example, suppose that Goetz\* possessed his racist belief or desire because he was the victim of prior attacks involving black assailants.<sup>99</sup> Indeed, suppose he despises himself for what he has become, and has tried without success to get 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 don't 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, but for who he is.<sup>100</sup> The target of the state's punishment continues to be his character, but a liberal state worthy of the name cannot take character to be a 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 BELIEF-CHOICE AND CHARACTER-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

---

<sup>98</sup> 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).

<sup>99</sup> See, e.g., Armour, *supra* note 51, at 799 (describing such a person as an "involuntary negrophobe"). Armour argues that such an actor should not be entitled to claim self-defense because "legal recognition of the Involuntary Negrophobe's claims would subvert the general welfare by destroying the legitimacy of the courts." *Id.* at 802.

<sup>100</sup> See, e.g., MOORE, *supra* note 22, 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.").

base the forfeiture on some choice he made that he should not have made, or on some choice he failed to make that he should have made. These theories therefore ground Goetz\*'s responsibility for believing that  $p$  in some choice or choices 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.

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 failed to do so. In other words, he chose to become or remain a racist when he should not have so chosen. The object of the forbidden choice are those acts or omissions that caused him to possess the belief or desire that  $q$ .<sup>101</sup>

---

<sup>101</sup> 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. He should have stopped his stereotypical beliefs 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, it is unlikely to be subject to one's conscious will. In other words, an actor is unlikely to be able consciously to control such self-control. See, e.g., John A. Bargh, *The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effect*, in *DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY* 361, 378 (Shelly Chaiken & Yaacov Trope eds., 1999) ("[The 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 Antisocial Prejudice to Behavior: The Moderating Role of Conscious Beliefs About Gender and Behavioral Control*, 91 J. PERSONALITY AND SOC. PSYCH. 268, 277 (2006) ("[T]he present data illustrate that relatively spontaneous interpersonal actions can be modified by motivation and control. . . . Future 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 339, 355 (Shelly Chaiken & Yaacov Trope eds., 1999) (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, *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 non-prejudiced 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 137, 145 (Stuart Oskamp ed., 2000) ("[S]elf-regulation, extended over time, may produce changes even in previously automatic, implicit negative responses."); Jack Glaser and John F. Kihlstrom, *Compensatory Automaticity: Unconscious Volition Is Not an Oxymoron*, in THE NEW UNCONSCIOUS 171, 171 (Ran R. Hassin 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 Cue for Control*, 83 J. PERSONALITY AND SOC. PSYCHOL. 1029, 1045 (2002) ("[P]eople can learn to put the brakes on their prejudices and control the influence of processes that otherwise could result in racially biased behavior"); Kerry Kawakami, *Just Say No (to Stereotyping): Effects of Training in the Negation of Stereotypic Associations on Stereotype Activation*, 78 J. PERSONALITY AND 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 Goals Prevent the Activation of Stereotypes*, 18 SOC. COGNITION 151, 173 (2000) ("Chronic [egalitarian] goals disrupt stereotype activation.").

This alternative theory is perhaps best understood as a variation on the character-choice theory. *See infra* pp. 49-57. 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

A. The Belief-Choice Theory

Recall Goetz\*'s psychology at the moment he pulled the trigger.<sup>102</sup> 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 danger, or he may not have. Either way, he formed the belief that it was (that  $p$ ). 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 killing his assailant was the only way 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.<sup>103</sup> 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 one can choose to do. Third, he is in control when he chooses to form the intent to kill.

---

at least chose not to try to develop, the right cognitive habits. 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 AND 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.”).

<sup>102</sup> See *supra* p. 29.

<sup>103</sup> 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 choose to think, but we cannot choose not to think, though we can choose to do things to try to distract ourselves from thinking.

Choosing, like thinking, is a mental act over which we have control.<sup>104</sup> 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.

But what's 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. 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*.<sup>105</sup> 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 also affect in turn 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 accept or reject

---

<sup>104</sup> See, e.g., ALFRED R. MELE, *MOTIVATION AND AGENCY* 198 (2003) (Practical deciding is a "momentary mental action of intention formation."); 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."). But see Pamela Hieronymi, *The Will as Reason 1* (Aug. 4, 2006) (unpublished manuscript, on file with author) ("In its practical employment, our capacity to reason is directed instead at the question of whether to  $\Phi$ ; it concludes, not in a judgment about  $\Phi$ -ing, but rather in an intention to  $\Phi$ .").

<sup>105</sup> See, e.g., DAVID OWENS, *REASON WITHOUT FREEDOM* 85 (2000) ("[B]elief is not subject to the will."); William 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."); 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."); Dion Scott-Kakures, *On Belief and Captivity of the Will*, 53 *PHIL. & PHENOMENOLOGICAL RES.* 77, 77 (1993) (arguing that it is conceptually, and not merely contingently, true that "with respect to our beliefs our wills are captive").

those beliefs.<sup>106</sup> 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.

If so, then the belief-choice theory is a non-starter. 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*,<sup>107</sup> just as he is

---

<sup>106</sup> 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 Belief in Criminal Law*, in CRIMINAL LAW THEORY 192 (Stephen Shute & A.P. Simister eds., 2002) (“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”).

<sup>107</sup> 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 63, 63 (Mattias Steup ed., 2001) (defending the “naive intuition that coming to believe something just by deciding to do so is possible”); Christoph Jäger, *Epistemic Deontology, Doxastic Voluntarism, and the Principle of Alternative Possibilities*, in KNOWLEDGE AND BELIEF 217, 226 (Winfried Löfer & 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].”); Mattias Steup, *Doxastic Voluntarism and Epistemic Deontology*, 15 ACT ANALYTICA 25, 26 (2000) (arguing that “[i]f we use the concept of [voluntary control] in its compatibilist sense, we get the result that we enjoy almost unconstrained voluntary control over our doxastic attitudes”). *But see* Nikolaj Nottelman, *The Anaological Argument for Doxastic Voluntarism*, 131 PHIL. STUD. 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”).

responsible for his character,<sup>108</sup> but whatever responsibility he bears for that belief is too weak to support punishing him for it. A liberal state cannot punish him 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 have chosen not to believe that *p*, but rather because he should have chosen not to possess the racist beliefs or desires causing him to form the belief that *p*. 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

---

Others argue that we bear some responsibility for our beliefs even though we do not have the same sort of control over them as we do over our actions. See, e.g., Adams, *supra* note 96, 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: ESSAYS ON EPISTEMIC JUSTIFICATION, RESPONSIBILITY, AND VIRTUE 93, 105 (Mattias Steup ed., 2001) (The conclusion that “neither believing nor forming beliefs is a case of action . . . [d]oes not prevent our sustaining a deontic version of an ethics of belief.”); Richard Feldman, *Voluntary Belief and Epistemic Evaluation*, in KNOWLEDGE, TRUTH, AND DUTY: ESSAYS ON EPISTEMIC JUSTIFICATION, RESPONSIBILITY, AND VIRTUE 77, 90 (Mattias Steup ed., 2001) (concluding that “deontological judgements about belief . . . do not imply that belief is voluntary.”); Pamela Hieronymi, *Responsibility for Believing 2* (May 3, 2006) (unpublished manuscript, on file with author) (“[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 THE MONIST 436, 436–37 (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 96, 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.”).

<sup>108</sup> See *supra* p. 41.

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 upon realizing that he is already in possession of the prohibited item, he fails to do something to dispossess himself of it.<sup>109</sup> The target of the state's punishment is not possession of the proscribed item itself. It is the choice to cause oneself to come into possession of it, or his choice to retain possession of it when he could and should have gotten rid of it.

### C. Punishing the Choice 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 that which triggers the forfeiture is a choice the actor makes, and not his character. 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, and 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.<sup>110</sup>

---

<sup>109</sup> See MODEL PENAL CODE § 2.01(4) (“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.”).

<sup>110</sup> Together with the problem of disproportionality associated with any forfeiture rule. See *supra* p. 23.

## 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 Klu Klux Klan, hoping that he would thereby become a racist, or at least aware of some prospect that he would become one. 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, at least insofar 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 efforts fail. Despite his regular attendance at rallies, cross-burnings and the like, he never ends up believing racist stereotypes or harboring racial animus. 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 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 become a racist. 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 luck.<sup>111</sup> Murderers are punished more than attempted murderers, even if luck is the only thing

---

<sup>111</sup> See, e.g., Ken Levy, *The Solution to the Problem of Outcome Luck: Why Harm Is Just as Punishable as the Wrongful Action that Causes It*, 24 *LAW & PHIL.* 263, 267-68 n.7 (2005) (collecting the latest literature on the question whether the extent of one's criminal liability should or should not depend on actually causing of the harm intended or risked).

that sets them apart. If you sneeze just as the trigger is pulled, thereby missing the target, then the crime is attempted murder. If no sneeze, and the target is killed, then 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 different fates of Goetz\* and Goetz\*\*. Perhaps. Yet even if their different fates is not a problem, or not much of one, the character-choice theory has two more.

## 2. Legality

The second problem with the character-choice theory is legality. Forfeiture rules in effect 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. Moreover, in the law's eyes 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.

No state makes it a crime to expose oneself to a threat. But a state could criminalize that choice if wanted to. If it wanted, a state could 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. The same goes for the intoxication forfeiture rule. If it wanted, a state could make it a crime to consume intoxicating substances, no matter what happens thereafter. These new crimes might be unwise, or even silly, but they would not be unfair, assuming one was aware of the new prohibitions. Compliance would not be difficult. Not drinking, or avoiding threatening situations, is 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 law-making power to prosecutors, jurors, and judges.<sup>112</sup> 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. An actor commits a crime — enforced via a forfeiture rule — when prosecutors, juries, and judges say 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

---

<sup>112</sup> Cf. 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).”).

don't consort with skinheads.

Of course, most of us don't need to join the KKK or hob-nob 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 (known as the IAT) and realize that we automatically associate black with bad.<sup>113</sup> The rest of us are strangers to ourselves, blissfully unaware, willfully ignorant, or self-deceived. In addition, most of us manage to become racists without trying. We don't need to *do* anything to acquire racist beliefs. We are not born with racist beliefs. We are taught them. For the unlucky, their parents, families, and friends are their first teachers. For the lucky, who have grown up among an enlightened circle of intimates, popular culture transmitted through television steps in to teach the association.<sup>114</sup> Picking up 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 still be not guilty if those beliefs

---

<sup>113</sup> See Project Implicit Home Page, <https://implicit.harvard.edu/implicit>. For the initial research "apprais[ing] the IAT's usefulness for measuring evaluative associations the underlie implicit attitudes," see Anthony G. Greenwald *et al.*, *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 J. PERSONALITY AND 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 AUTOMATIC PROCESSES IN SOCIAL THINKING AND BEHAVIOR (John A. Bargh ed., forthcoming) (Feb. 3, 2005) (manuscript at 3); *But see* Hal R. Arkes & Philip E. Tetlock, *Attributions of Implicit Prejudice, or "Would Jesse Jackson 'Fail' the Implicit Association Test?"*, 15 PSYCH. INQ. 257, 257 (2004) (offering "three objections to the inferential leap from the comparative [reaction time] of different associations to the attribution of implicit prejudice").

<sup>114</sup> 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[.]").

were tucked away in unconscious. Given the subtle nature of modern-day racism, this description probably applies to many, of 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,<sup>115</sup> even if he need not be aware of the obligation itself.<sup>116</sup>

Maybe the legislature should instead demand that all adults periodically attend state-run 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 need to do is attend class. Consequently, the principle of legality would not stand in the way of the state creating such a crime. But it seems to me that another principle would.

---

<sup>115</sup> See, e.g., Larry Alexander, *Criminal Liability for Omissions: An Inventory of Issues*, in *CRIMINAL LAW THEORY: DOCTRINES OF THE GENERAL PART* 121, 124 (2002) (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”).

<sup>116</sup> 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).

### 3. Liberalism

The third problem with the character-choice theory is the hardest to avoid. According to the prevailing orthodoxy,<sup>117</sup> the only reason for which a liberal state can make it a crime to do or not do something is to prevent harm, even if the harm targeted is quite remote.<sup>118</sup> According to J.S. Mill's influential formulation of this "harm principle": "[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."<sup>119</sup> It cannot legitimately criminalize thought and beliefs.<sup>120</sup> Presumably, therefore, liberal states also lack the authority to force a

---

<sup>117</sup> 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*").

<sup>118</sup> See, e.g., Andrew von Hirsch, *Extending the Harm Principle: 'Remote' Harms and Fair Imputation*, in *HARM AND CULPABILITY* 259, 276 (A.P. Simester & A.T.H. Smith eds., 1996) (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).

<sup>119</sup> 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 *OXFORD J. LEG. 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).

<sup>120</sup> See, e.g., Shlomit Wallerstein, *Criminalizing Remote Harm and the Case of Anti-democratic Activity*, 28 *CARDOZO L. REV.* (forthcoming 2007) (manuscript at 4, on file with author) ("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.")

person on pain of punishment to do or not to do things with the purpose of preventing them from forming state-disapproved beliefs or desires, or with the purpose of causing them to shed such beliefs or desires.

A liberal state is free to combat racism in a number of ways. It is free to coerce children to attend school, and to inculcate, or if you prefer, indoctrinate them in the virtues of liberal citizenship, virtues which have no place for the vice of racism. A liberal state is likewise free to try to persuade its adult citizens to reject racism, speaking out loud and clear against it. A liberal state might even be free to condition the availability of certain benefits — such as the right to carry a firearm — on a citizen’s willingness to take part in programs designed to rid participants of racist beliefs. Indeed, more controversially, it might even be free to do things *to* its citizens — like expose them to various debiasing stimuli<sup>121</sup> — in order to break the implicit association between black and bad.

But again, what a liberal state cannot do is to force its citizens on pain of punishment to do things designed to rid them of their racism, or to prevent 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 those beliefs and desires may be, whether the punishing is done directly through rules of liability, or indirectly through forfeiture rules.

#### CONCLUSION

Cases like that of Bernhard Goetz, or at least like that of Goetz\*, leave the liberal state in a tough position. On the one hand, it

---

<sup>121</sup> See Kang, *supra* note 114, 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”).

understandably wants not to be seen as condoning the racism that causes a citizen like Goetz\* to believe that he is about to be killed. On the other hand, the theories examined here, theories upon which Goetz\*'s might lose his claim of self-defense, are theories unavailable to a liberal state. 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 being who they are, nor for choosing to become who they are, or for remaining who they are.

If an actor kills only because he believed he was about to be killed, and if he believed he was about to be killed only because he was a racist, we can and should condemn the racism that lead 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 should forfeit his claim of self-defense. Punishing an actor like Goetz\* is not the liberal way to get to a liberal society. Quite the contrary: foregoing the punishment of such an actor is the price one pays for a society in which the only legitimate basis upon which a citizen can be punished is that he has chosen to cause or risk causing harm when the law does not permit him to make such a choice.