9-13-2006

What's Wrong with Involuntary Manslaughter?

Stephen P. Garvey
Cornell Law School, spg3@cornell.edu

What's Wrong with Involuntary Manslaughter?

Stephen P. Garvey*

Efforts to explain when and why the state can legitimately impose retributive punishment on an actor who inadvertently creates an unjustified risk of causing death (and death results) typically rely on one of two theories. The prior-choice theory claims that retributive punishment for inadvertent lethal risk creation is justified if and only if the actor's inadvertence or ignorance was a but-for and proximate result of a prior culpable choice. The hypothetical-choice theory claims that retributive punishment for inadvertent lethal risk creation is justified if and only if the actor would have chosen to take the risk if he had been aware of it, even though he was not in fact aware of it.

I argue that neither of these theories satisfactorily identifies when and why retributive punishment is warranted for inadvertent lethal risk creation. Instead, I propose that an actor who creates a risk of causing death, but who was unaware of that risk, can fairly be subject to retributive punishment if he was either nonwillfully ignorant or self-deceived with respect to the existence of the risk, and if such ignorance or self-deception was due to the causal influence of a desire he could and should have controlled. The culpability of such an actor consists, not in any prior actual choice to do wrong, nor in any imagined hypothetical choice to do wrong, but in the culpable failure to exercise doxastic self-control: control over one's beliefs.

Walter and Bernice Williams were loving parents. They were also convicted of letting their fourteen-month-old son die of pneumonia.[1] The Williamses said they never realized their son's life was in danger, and indeed, the state of Washington never claimed otherwise. It wanted to punish them anyway. But if the Williamses never realized their son's life was in danger, how can they fairly be punished for endangering it? If we punish them nonetheless, why do we punish them? What do we punish them for?

---

* Professor of Law, Cornell Law School. Thanks to Robert Blecker, John Blume, Steven Clymer, Valerie Hans, James Henderson, Kenneth Simons, Sheri Johnson, and Trevor Morrison for helpful comments.

1. State v. Williams, 484 P.2d 1167, 1174 (Wash. Ct. App. 1971). The case is reproduced in a number of criminal law casebooks. *See, e.g.*, KATE E. BLOCH & KEVIN C. MCMUNIGAL, CRIMINAL LAW 392–95 (2005); JOSHUA DRESSLER, CASES AND MATERIALS ON CRIMINAL LAW 296–99 (3d ed. 2003); SANFORD H. KADISH & STEPHEN J. SCHULHOFER, CRIMINAL LAW AND ITS PROCESSES 445–48 (6th ed. 1995); JOHN KAPLAN, ROBERT WEISBERG & GUYORA BINDER, CRIMINAL LAW 370–73 (5th ed. 2004); CYNTHIA LEE & ANGELA HARRIS, CRIMINAL LAW 425–29 (2005); LLOYD L. WEINREB, CRIMINAL LAW 290–91 (7th ed. 2003). For a fictionalized account offering a similar fact pattern, see Norval Morris, *The Watching Brief*, 54 U. CHI. L. REV. 1215 (1987).

Walter Williams was described as a "24-year-old full-blooded Sheshont Indian with a sixth-grade education."[2] Bernice was described as a "20-year-old part Indian with an 11th grade education."[3] In the fall of 1968, their infant son became ill.[4] Walter and Bernice realized he was ill but thought he had a simple toothache.[5] They gave him aspirin, believing that was all they needed to do.[6]

The boy was in truth suffering from "an abscessed tooth," which "develop[ed] into an infection of the mouth and cheeks, eventually becoming gangrenous."[7] The infection emitted the odor characteristic of gangrene,[8] and the boy's cheek "turned 'a bluish color like.'"[9] Unable to eat, the child became malnourished and eventually died of pneumonia.[10] Medical testimony established that he would have survived had he received adequate medical care at least one week prior to his death.[11] But he never received such care.[12] His parents failed to secure it for him.

The Williamses were physically and financially able to take the boy to a doctor.[13] In fact, Walter had taken him to a doctor for medical attention earlier in the year.[14] Yet on this occasion they did not seek help.[15] Believing the child was suffering from nothing but a toothache, Walter testified that given "the way the cheek looked, . . . and that stuff on his hair, [the welfare authorities] would think we were neglecting him and take him away from us and not give him back."[16] Bernice echoed her husband's fear,[17] and indeed, they had good reason to be afraid. Welfare authorities were at the time quick

2. *Williams*, 484 P.2d at 1169. For present purposes I assume that the Williamses had the cognitive capacity to realize their son's life was in danger, despite their limited education. In other words, I assume their limited cognitive abilities were not so severe as to have rendered them incapable of realizing their son's life was in danger based on the evidence available to them at the time they should have acted but failed to do so.

3. *Id.* at 1170.

4. *Id.* The child's name was Walter Joseph Tabafunda. *Id.* Walter Williams was not the child's biological father, nor had he adopted the child. *Id.* at 1172. According to the appellate court, however, the "evidence show[ed] that he had assumed responsibility with his wife for the care and maintenance of the child, whom he greatly loved." *Id.*

5. *Id.* at 1170.

6. *Id.* at 1174.

7. *Id.* at 1173.

8. *Id.*

9. *Id.* at 1174.

10. *Id.* at 1173.

11. *Id.* at 1173–74.

12. *Id.* at 1170.

13. *Id.* at 1174.

14. *Id.*

15. *Id.* at 1170.

16. *Id.* at 1174 (internal quotations omitted).

17. *Id.*

to remove Native American children from their families and place them with non-Native American families.[18]

Charged with what is usually described as involuntary manslaughter[19] or negligent homicide,[20] the trial court expressly found that the Williamses "did not realize how sick the baby was."[21] Moreover, the court found that the Williamses "loved the baby."[22] Nonetheless, their failure to realize their son's life was in danger when a "man of reasonable prudence under the same or similar conditions"[23] would have realized the risk, coupled with their failure to take the baby to a doctor when they were obligated to do so, added up

18. *See* William Byler, *Removing Children: The Destruction of American Indian Families*, C.R. DIG., Summer 1977, at 19, 19 ("In the State of Washington, the Indian adoption rate is 19 times greater and the foster care rate 10 times greater [than for non-Indian children].").

19. Washington law at the time of the *Williams* case denominated this offense "manslaughter." *See* State v. Williams, 484 P.2d 1167, 1171 & n.2 (Wash. Ct. App. 1971). The appellate court emphasized that manslaughter under Washington law required only "simple" or "ordinary" negligence, in contrast to "gross" or "criminal" negligence associated with involuntary manslaughter at common law. *Id.* at 1171. Consequently, the case might have come out differently if Washington law had required gross or criminal negligence in order to sustain a conviction for manslaughter. The distinction between civil and criminal negligence is in any event an elusive one. *See, e.g.*, MODEL PENAL CODE § 210.4 cmt. 2, at 83 (1980) ("While some of the earliest negligent homicide statutes required only ordinary negligence for criminal liability, the trend has been to require something more, though . . . the 'something more' typically was not described with precision. The same is true of judicial decisions on the subject."); WAYNE R. LAFAVE, CRIMINAL LAW § 3.7(b), at 249 (3d ed. 2000) ("Though the legislatures and the courts have often made it clear that criminal liability generally requires more fault than the ordinary negligence which will do for tort liability, they have not so often made it plain just what is required in addition to tort negligence . . . ."). The Model Penal Code provides that an actor's failure to perceive a risk does not rise to the level of culpable negligence unless the actor's failure amounts to a "gross deviation from the standard of care that a reasonable person would observe in the actor's situation." MODEL PENAL CODE § 2.02(2)(d) (1962). According to some commentators, this gross-deviation requirement embodies the Code's effort to distinguish criminal from civil negligence. *See, e.g.*, MARKUS D. DUBBER, CRIMINAL LAW: MODEL PENAL CODE § 4.2(C)(iv), at 78 n.88 (2002); Peter W. Low, *The Model Penal Code, the Common Law, and Mistakes of Fact: Recklessness, Negligence, or Strict Liability?*, 19 RUTGERS L.J. 539, 545–46 (1988).

20. The Model Penal Code does not use the term "involuntary manslaughter." It distinguishes instead between "manslaughter," which involves the conscious disregard of a substantial and unjustifiable risk of causing death (with death resulting), and "negligent homicide," which involves the failure to perceive a substantial and unjustifiable risk of causing death (with death resulting). *See* MODEL PENAL CODE § 210.3(1)(a) (1962) (defining "manslaughter"); *id.* § 210.4 (defining "negligent homicide"). The common law drew no distinction between manslaughter and negligent homicide so defined, treating both as cases of "involuntary manslaughter." *See* MODEL PENAL CODE § 210.4 cmt. 1, at 81 (1980) ("[I]nvoluntary manslaughter [at common law] was a catch-all offense that did not distinguish between recklessness and negligence . . . ."); *cf.* ROLLIN M. PERKINS & RONALD N. BOYCE, CRIMINAL LAW 107 (3d ed. 1982) (stating that awareness of risk was an element to be considered in determining if negligence was criminal but further stating that "[w]hether [at common law] negligence [was] criminal or ordinary . . . depend[ed] not upon the element of awareness but upon the degree of the negligence").

21. *Williams*, 484 P.2d at 1170.

22. *Id.*

23. *Id.* at 1174.

to involuntary manslaughter.[24] The Williamses were sentenced to three years imprisonment, with their sentences suspended.[25]

The case can elicit a range of responses.[26] On the one hand, you might think that the Williamses did realize that failing to take their son to a doctor would jeopardize his life, despite their claims to the contrary. You might also think that the Williamses had no good reason to take that risk. If so, then the Williamses were guilty of reckless homicide, not simply involuntary manslaughter. They took a risk, and the risk they took was one the law did not permit them to take. Alternatively, you might think that the Williamses were justified in risking their son's life, since going to the doctor meant taking a risk of losing him to the welfare authorities. The law should have allowed the Williamses to risk the child's life in order to avoid the risk of having him taken away. If so, then the Williamses should have been acquitted. They did nothing the law should forbid.

On the other hand, you might think that the Williamses were telling the truth when they said that they did not realize their son's life was in danger. You might nonetheless also think that the Williamses should have seen the risk because a reasonable person in their situation would have seen it. If so, then the actual outcome of the case was the right one. The Williamses were guilty of involuntary manslaughter. Alternatively, you might think that a reasonable person in the Williamses' situation would not have realized their son's life was in danger. If so, then the Williamses should have been acquitted. They believed as a reasonable person would have believed, and the criminal law has no business punishing them.

All these reactions take for granted that the state can, at least under some circumstances, legitimately punish inadvertent risk creation. But according to one prominent school of thought, sometimes known as choice retribution,[27] this assumption is mistaken. On this view an actor can legitimately be punished only if he deserves to be punished, and an actor deserves to be punished only if he has freely chosen to act in a way the state demands he refrain from acting or not to act in a way it demands he act.[28] If the

---

24. *Id.*

25. PAUL H. ROBINSON, WOULD YOU CONVICT? SEVENTEEN CASES THAT CHALLENGED THE LAW 142 (1999).

26. *See id.* at 140–41 (reporting results of a survey based on the facts of the *Williams* case finding that "[t]hirty-seven percent of the people would impose no punishment, but 45 percent would impose a substantial punishment of a year or more in prison").

27. Theorists typically contrast choice retribution with character retribution. Choice retribution holds that the punishment an actor deserves depends on the choices he has made. Character retribution holds that the punishment an actor deserves depends on the character he has revealed through the choices he has made. *See infra* notes 66–68 and accompanying text.

28. *See infra* note 68 and accompanying text (discussing choice retribution). Although I will hereafter usually speak in terms of an actor choosing to act so as to create a lethal risk, this way of speaking should be understood to include *mutatis mutandis* cases in which an actor chooses not to act when he is under an obligation to act, and thereby fails to extinguish such a risk, as in the *Williams* case itself.

pertinent demand is to refrain from imposing on others an unjustified risk of causing death, how can an actor who sees no such risk in what he is doing be said to have chosen to take that risk and thus to have chosen to act in a way the law demands he refrain from acting? A choice retributivist would therefore seem to have no choice: The Williamses must be let go.

Nonetheless, at least some cases involving the inadvertent creation of a lethal risk do seem to evoke a retributive response, even if the basis for such a response is obscure. Take Sam and Tiffany.[29] Like Walter and Bernice, their child is very ill. Unlike Walter and Bernice, Sam and Tiffany are well-educated and well-to-do. They are also single-minded social climbers. Imagine further that Sam and Tiffany have some inkling that their child's illness poses a threat to his life,[30] but they never actually come consciously to believe that his life is in danger. Imagine finally that the reason they fail to form that belief is because they have become preoccupied with planning the "party of the decade."[31] Failing to see the risk to their child's life, they do nothing to dissipate it, and the child dies.

Different people will react differently to the facts surrounding the death of Sam and Tiffany's child, just as they do to the facts surrounding the death of Walter and Bernice's. Some might believe neither couple deserves retributive punishment; others might believe both do. But what if your intuition tells you to punish Sam and Tiffany but not Walter and Bernice? Does a normatively defensible distinction exist between the two cases? In other words, if some cases of involuntary manslaughter warrant retributive punishment, but no cases of involuntary manslaughter involve a conscious choice to risk causing death, then when, if at all, is retributive punishment for the inadvertent creation of a lethal risk warranted, and why?

The answer I propose is this: An actor who creates a risk of causing death but who was unaware of that risk is fairly subject to retributive punishment if he was either *nonwillfully ignorant* or *self-deceived* with respect to the existence of the risk, and if such ignorance or self-deception was due to the causal influence of a desire he should have controlled. The culpability of such an actor does not consist in any choice to do wrong, but rather in the culpable failure to exereise *doxastic self-control*, i.e., control over desires that influence the formation and awareness of one's beliefs. An actor who is nonwillfully ignorant allows desire to preclude him from forming the belief that he is imposing a risk of death when the evidence available to him

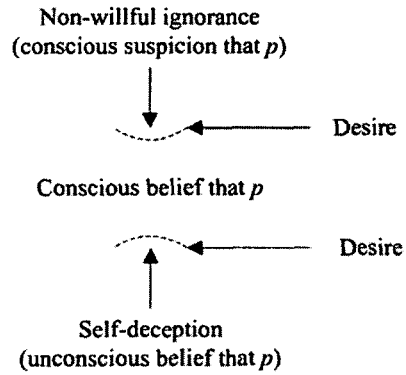
29. This hypothetical is based on DRESSLER, *supra* note 1, at 301–02, which is based in turn on the hypothetical in Larry Alexander, *Reconsidering the Relationship Among Voluntary Acts, Strict Liability, and Negligence in Criminal Law*, 7 SOC. PHIL. & POL'Y 84, 100 (1990) [hereinafter Alexander, *Reconsidering the Relationship*]. See Larry Alexander, *Insufficient Concern: A Unified Conception of Criminal Culpability*, 88 CAL. L. REV. 931, 950 (2000) [hereinafter Alexander, *Insufficient Concern*], for a reprised version of his original hypothetical.

30. As explained later, this "inkling" would more precisely be described either as a suspicion or as an unconscious belief that the child's life was in danger.

31. DRESSLER, *supra* note 1, at 301.

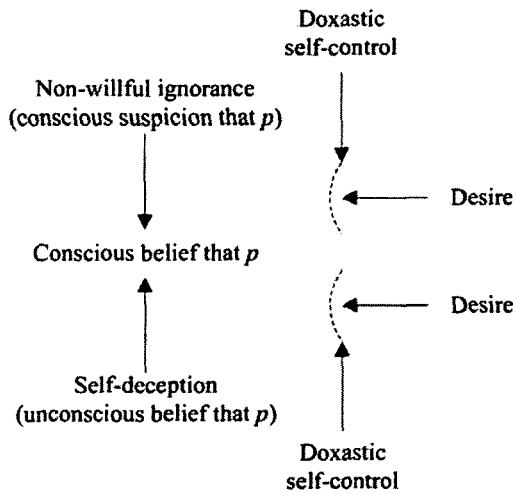
supports the formation of that belief, while an actor who is self-deceived forms that belief but allows desire to prevent him from becoming aware of it. In either case the actor could and should have controlled the wayward desire, thereby allowing the relevant belief to form and surface into awareness.

Before the actor's exercise of doxastic self-control, his cognitive state can be depicted as follows, where the belief that *p* is the actor's belief that his conduct is creating a lethal risk.



Before the exercise of doxastic self-control

Desire prevents the formation of the conscious belief that the actor is creating a lethal risk. In contrast, the actor's cognitive state after the exercise of doxastic self-control might be depicted as follows, where the exercise of doxastic self-control cuts off the influence of desire and allows the actor to form the conscious belief that he is creating a lethal risk.



After the exercise of doxastic self-control

Thus the fact—assuming it to be a fact—that the Williamses never consciously believed their son's life was in danger is not enough to inoculate them against retributive punishment. On one plausible reading of the facts, the Williamses either suspected their son's life was in danger, though they did not believe it was, or they believed his life was in danger, though they remained unaware of that belief. They were, in other words, either nonwillfully ignorant or self-deceived. In either case the cognitive state in which they found themselves was the result of desire. They did not want consciously to believe that their inaction presented any risk to their son's life, because if they did so believe, they would have had to choose between risking the child's life and risking his loss to the state. They did not want to face that choice. Nonetheless, their ignorance was at least arguably excusable, for the desire blinding them to the truth was the desire not to lose a child they dearly loved.

In contrast, excusing Sam and Tiffany would be misplaced. If we assume that they too were either nonwillfully ignorant or self-deceived with respect to the existence of the risk to their child's life, then like the Williamses, they should have exercised doxastic self-control so as to form the conscious belief that their child's life was in danger. Yet unlike the Williamses, the desire blinding them to the truth was not one the law should excuse. The desire not to lose one's child is one thing. The desire to move one step higher on the social ladder is another. Yet if Sam and Tiffany are punished, they are punished not because they possessed that desire. They are punished because they failed to control the influence of that desire on their ability to form the conscious belief that their son's life was in jeopardy.

The argument unfolds in three Parts. Part I identifies as precisely as possible what it is that distinguishes involuntary manslaughter from reckless homicide. Part II explores two theories that attempt to reconcile liability for involuntary manslaughter with retribution's commitment to liability based on choice. Neither provides a satisfactory way to separate cases in which punishment for the inadvertent creation of a lethal risk is warranted from those in which it is not. Part III argues that retributive punishment for such unwitting risk creation is warranted if and because the actor's failure to perceive the lethal risk is due to a culpable failure to exercise doxastic self-control.

I. Involuntary Manslaughter and Reckless Homicide

According to the Model Penal Code, involuntary manslaughter (negligent homicide in the Code's terminology[32]) and reckless homicide (manslaughter in the Code's terminology[33]) both involve an actor who

32. MODEL PENAL CODE § 210.4 (1962).
33. *Id.* § 210.3.

creates an unjustified risk of causing death.[34] But in the Code's hierarchy of homicide offenses, involuntary manslaughter is a lesser offense deserving of less punishment.[35] So what distinguishes involuntary manslaughter from reckless homicide, making the former deserving of less punishment than the latter?

According to the Code, an actor is guilty of reckless homicide when he "recklessly" causes the death of another human being.[36] An actor is "reckless" with respect to a result element of an offense, such as the death of another human being, "when he consciously disregards a substantial and unjustifiable risk that [death] . . . will result from his conduct."[37] Moreover, the "risk must be of such a nature and degree that, considering the nature and purpose of the actor's conduct and the circumstances known to him, its disregard involves a gross deviation from the standard of conduct that a law-abiding person would observe in the actor's situation."[38]

The Code's definition of reckless homicide, which incorporates its definition of recklessness, is no model of clarity.[39] The questions it raises are many. For example, what is the relationship between the requirement that the risk be "substantial and unjustifiable," and that which requires the risk to be such that its disregard involves a "gross deviation" from the "standard of conduct that a law-abiding person would observe in the actor's situation"? Does the second clause define, or perhaps refine, in what a "substantial and unjustifiable" risk consists,[40] or does it state a separate and independent requirement?[41] In either case, must the risk be *both* substantial *and*

---

34. *See* MODEL PENAL CODE § 210.4 cmt. 1, at 81 (1980) ("[T]he reckless actor must 'consciously disregard' a substantial and unjustifiable homicidal risk created by his conduct, whereas the negligent actor need only disregard a risk of which he 'should be aware.'"); *id.* § 210.3 cmt. 3 at 51 ("[T]here must be a substantial and unjustifiable risk of homicide to establish recklessness . . . .").

35. The Code grades negligent homicide as a third degree felony, *id.* § 210.4(2), and reckless homicide as a second degree felony. *Id.* § 210.3(2).

36. *Id.* § 210.3(1)(a).

37. *Id.* § 2.02(2)(c). For a helpful comparison between the Model Penal Code definition of recklessness and the various then-extant state law definitions of recklessness, see David M. Treiman, *Recklessness and the Model Penal Code*, 9 AM. J. CRIM. L. 281 app. II, at 376 (1981).

38. *See* MODEL PENAL CODE § 2.02(2)(c) (1962). English law distinguishes between "advertent recklessness" (also known as "*Cunningham* recklessness") and "inadvertent recklessness" (also known as "*Caldwell* recklessness"). *See, e.g.*, ANDREW ASHWORTH, PRINCIPLES OF CRIMINAL LAW § 5.3(c), at 180–87 (3d ed. 2003).

39. The most thorough analysis of the Code's definition of reckless of which I am aware is Treiman, *supra* note 37. See Paul Robinson, *Prohibited Risks and Culpable Disregard or Inattentiveness: Challenge and Confusion in the Formulation of Risk-Creation Offenses*, 4 THEORETICAL INQUIRIES L. 367, 371–77 (2003), for a helpful exegesis of the Code's definition of recklessness, and Kenneth W. Simons, *Should the Model Penal Code's* Mens Rea *Provisions Be Amended?*, 1 OHIO ST. J. CRIM. L. 179, 188–95 (2003), for a description of some of the problems with the Code's definition of recklessness.

40. *See, e.g.*, Treiman, *supra* note 37, at 329 (suggesting this interpretation as one possibility).

41. *See, e.g., id.* at 334 (concluding that the Code intended this interpretation but recognizing that the language of the provision itself is ambiguous). Another interpretation of the relationship between the two requirements would hold that the first asks whether the actor's reasons for taking

unjustifiable, as the text of the definition alone would suggest,[42] or is the substantiality of a risk nothing more than a factor to be considered in assessing the risk's unjustifiablity, such that the risk's unjustifiability is really all that matters in the end?[43] If substantiality and unjustifiability are separate elements, what makes a risk substantial?[44]

Finally, what exactly is it that the actor must "consciously disregard"? Will disregard of *any* risk suffice?[45] Or must the actor also disregard the risk's substantiality, whatever "substantial" turns out to mean?[46] Or indeed, must the actor disregard not only the risk's substantiality but also the fact that the law considers the taking of it unjustifiable?[47] In all events, what does it

---

the risk render his taking of it justifiable or permissible, while the second asks whether the actor's reasons for taking the risk, though not rendering its taking justifiable or permissible, nonetheless render its taking excusable. *See* GEORGE P. FLETCHER, BASIC CONCEPTS OF CRIMINAL LAW 126 (1998).

42. *See* MODEL PENAL CODE § 2.02(2)(c) (1962) ("A person acts recklessly with respect to a material element of an offense when he consciously disregards a substantial and unjustifiable risk that the material element exists or will result from his conduct.").

43. *See, e.g.*, Alexander, *Insufficient Concern*, *supra* note 29, at 934 ("[R]ecklessness consists of imposing unjustifiable risks on others. The level of risk imposed will bear on its justifiability but is not itself an independent criterion of recklessness.").

44. *See, e.g.*, Treiman, *supra* note 37, at 337–38 (describing two possible interpretations of the substantiality requirement); *see also* Joshua Dressler, *Does One Mens Rea Fit All?: Thoughts on Alexander's Unified Conception of Criminal Culpability*, 88 CAL. L. REV. 955, 959 (2000) (arguing that "substantial" should be read to modify "unjustifiable," such that the reckless actor is one who believes he is taking a "substantially unjustified" risk). Another interpretation of the substantiality requirement, akin to that of Dressler's, maintains that the requirement's function is to preclude liability for the taking of a risk that, though unjustified, is nonetheless insubstantial inasmuch as the taking of it would involve a "de minimis infraction[]" in the sense provided for in the Code. *See* Treiman, *supra* note 37, at 337–38 ("The second function that the requirement of substantiality might serve is as an exclusion of de minimis violations of the law."). The Code describes the circumstances under which a court "shall dismiss a prosecution" if and because it finds the defendant's conduct involves a de minimis infraction of the law in § 2.12. *See* MODEL PENAL CODE § 2.12 (1962).

45. *See, e.g.*, Treiman, *supra* note 37, at 361 (suggesting this interpretation as one possibility).

46. *See* Claire Finkelstein, *Responsibility for Unintended Consequences*, 2 OHIO ST. J. CRIM. L. 579, 594–95 (2005) (arguing that the reckless actor must be aware of the risk's substantiality); Kenneth W. Simons, *Culpability and Retributive Theory: The Problem of Criminal Negligence*, 5 J. CONTEMP. LEGAL ISSUES 365, 383 n.48 (1994) [hereinafter Simons, *Culpability and Retributive Theory*] (same); Kenneth W. Simons, *Does Punishment for "Culpable Indifference" Simply Punish for "Bad Character?,"* 6 BUFF. CRIM. L. REV. 219, 226 n.11 (2002) [hereinafter Simons, *Bad Character*] (stating that the Model Penal Code's "cognitive criterion of recklessness . . . require[s], inter alia, that the actor subjectively believe that she is creating a substantial risk"); Treiman, *supra* note 37, at 362 ("There is strong support for the view that the actor must be aware of a substantial risk, not just any risk.").

47. *Compare* Alexander, *Insufficient Concern*, *supra* note 29, at 937 (awareness of risk's unjustifiability not required), *and* Treiman, *supra* note 37, at 362 (arguing that the "actor need not be aware that the risk is unjustifiable"), *with* Dressler, *supra* note 44, at 959 ("In my view, what should justify a judgment of criminal recklessness is that the actor is aware that she is taking an unjustifiable risk or, at the very least, is aware that there is a significant likelihood that her intended conduct is unjustifiable . . . ."), *and* Stephen Shute, *Knowledge and Belief in the Criminal Law*, *in* CRIMINAL LAW THEORY 171, 182 (Stephen Shute & A.P. Simester eds., 2002) ("Subjectivists . . . [should] extend the requirement for knowledge or belief to the *actus reus* element that the risk is unjustified.").

mean to "consciously disregard" anything, whatever it is that one must so disregard?[48]

For present purposes I must set these difficult interpretive questions to one side. Nonetheless, because an analysis of involuntary manslaughter cannot intelligibly proceed without some analysis of reckless homicide in place, the analysis on which I will rely begins as follows. An actor is not guilty of reckless homicide unless, based on the beliefs available to him at the time he acts (his background beliefs), he forms the further belief that his act is creating a substantial risk of causing death (which belief I will hereafter refer to as the belief that *p*), keeping in mind that even a very small chance of causing death can fairly be characterized as "substantial." An actor is *not* reckless

---

Insofar the unjustifiability or impermissibility of the risk an actor takes is a question of law, those who maintain that an otherwise reckless actor need not be aware of the fact that the law judges his risk taking as unjustified or impermissible may do so on the ground that ignorance of the law is no excuse. But many scholars have persuasively argued that ignorance of the law, or at least reasonable ignorance of the law, should excuse, even if it now does not. *See, e.g.*, Douglas Husak & Andrew von Hirsch, *Culpability and Mistake of Law*, *in* ACTION AND VALUE IN CRIMINAL LAW 157, 173 (Stephen Shute et al. eds., 1993) ("[A] defendant who is ignorant of the applicable legal rule . . . should ordinarily . . . be excused if his legal mistake was a reasonable one . . . ; or . . . have his punishment mitigated if he did not know but should have known of the conduct's illegality."); Joshua Dressler, *Reflections on Excusing Wrongdoers: Moral Theory, New Excuses and the Model Penal Code*, 19 RUTGERS L.J. 671, 708 (1988) ("[A]n excuse defense should be allowed whenever a jury is satisfied that an actor's lack of knowledge or understanding of a penal law was a reasonable one . . . ."). Consequently, according to this line of thought an actor should only be described as reckless if he realized that the risk he was taking was legally unjustified or impermissible. Indeed, on this view the fact that an actor is unaware of the risk he is taking is ultimately important because and insofar that it causes him to be unaware of the fact that the risk he is taking is a risk the law demands he not take. *Cf.* Michael Zimmerman, *Moral Responsibility and Ignorance*, 107 ETHICS 410, 412 (1997) ("The ignorance that is directly relevant is [the actor's own] ignorance of his doing something . . . wrong.").

48. Treiman, *supra* note 37, at 369–71 (noting the ambiguity of "conscious disregard"). Markus Dubber suggests that "conscious disregard" might be construed to require the actor to "*accept*[ ] . . . the risk actually manifesting itself." *See* DUBBER, *supra* note 19, § 4.2(C)(iii), at 75–76. Consequently, Dubber argues that an actor who takes an unjustified risk believing it to be less-than-practically-certain to result should only be considered reckless if he accepts the risk "actually manifesting itself," whereas an otherwise similarly-situated actor who does not so accept the risk should presumably be considered merely negligent. *See id.* § 4.2(C)(iii), at 74–76. In contrast, Alan Michaels argues that an actor who takes an unjustified risk believing it to be less-than-practically-certain to result should be considered reckless, whereas an otherwise similarly-situated actor who accepts the risk should be treated as if he believed the risk were practically certain to result. *See* Alan Michaels, *Acceptance: The Missing Mental State*, 71 S. CAL. L. REV. 953, 955, 965–76 (1998). In other words, whereas Dubber suggests treating the accepting reckless actor as "reckless" and the nonaccepting reckless actor as "negligent," Michaels would treat the accepting reckless actor as "knowing" and the nonaccepting reckless actor as "reckless." Moreover, Dubber and Michaels each seem to suggest that his position finds support in the German law concept of *bedingter Vorsatz*, which is derived in turn from the Roman law concept of *dolus eventualis*. *See* DUBBER, *supra* note 19, § 4.2(C)(iii), at 74; Michaels, *supra*, at 1024–28. See Greg Taylor, *Concepts of Intention in German Law*, 24 OXFORD J. LEGAL STUD. 99, 116 (2004), for an in-depth examination of *dolus eventualis* in contemporary German law, which ultimately concludes that the "concept . . . is . . . irretrievably flawed," and Greg Taylor, *The Intention Debate in German Law*, 17 RATIO JURIS 346, 348–78 (2004), for a detailed discussion of the debate over *dolus eventualis* in the German criminal law literature.

simply because he *could* have formed the belief that *p* based on the beliefs available to him at the time. The actor must actually form the belief that *p*.

Importantly, this analysis presupposes that the reckless actor's belief that *p* possesses two key features. First, the belief must be a *belief*, and not merely a *suspicion*.[49] Or to put the point another way, the actor's belief that *p* must be a *full* belief, and not merely a *partial* one. For a belief to qualify as a (full) belief, and not merely as a suspicion (or partial belief), the actor must hold that belief with some appropriate measure of confidence and lack of doubt. If an actor lacks such confidence, or harbors such doubt, we would not say that he believes that *p*. We would instead say that he suspects that *p*, or something along those lines. Confidence and doubt are of course matters of degree.[50] The point at which an actor's suspicion (or partial belief) that *p* ripens into a (full) belief that *p* is elusive, but some such point exists all the same.

Second, the actor's belief must be a belief of which he is sufficiently aware or conscious. Consciousness or awareness, like confidence and doubt, are matters of degree.[51] For present purposes it suffices to distinguish two degrees or levels of awareness and two degrees or levels of unawareness: reflective awareness, dispositional awareness, modest unawareness, and deep unawareness. The first two of these states are states of *awareness*. An actor is *reflectively aware* of his belief that *p* if and when that belief is at the forefront of, or present in, the actor's mind. Reflective awareness might also be

49. The distinction at issue here can be described in various ways. *See, e.g.*, JONATHAN ADLER, BELIEF'S OWN ETHICS 231 (2002) (describing the distinction in terms of full belief versus partial or degrees of belief); MICHAEL E. BRATMAN, INTENTION, PLANS, AND PRACTICAL REASON 36–37 (1999) (describing the distinction in terms of flat-out belief versus degrees of confidence); RICHARD FOLEY, WORKING WITHOUT A NET 141 (1993) (describing the distinction in terms of degrees of belief); Shute, *supra* note 47, at 193 (describing the distinction in terms of belief versus suspicion); RICHARD SWINBURNE, EPISTEMIC JUSTIFICATION 34–35 (2000) (describing the distinction in terms of strong beliefs versus weak beliefs); ALAN R. WHITE, MISLEADING CASES 132 (1991) (describing the distinction in terms of belief versus suspicion). An actor can harbor some doubt about the truth of a proposition *p* without thereby losing his belief that *p*, but at some point his doubt must undermine his belief. Or to put the same point another way, while an actor can suspect that not-*p* without thereby losing his belief that *p*, at some point his suspicion that not-*p* will become large enough to undermine his belief that *p*. Some scholars would perhaps describe an actor who suspects that not-*p* while nonetheless believing that *p* as having made a "reckless mistake" with respect to *p* when not-*p* is true. *See* PAUL H. ROBINSON, CRIMINAL LAW 263 (1997) ("A *reckless mistake* occurs when an actor is aware of a substantial risk that the required circumstance exists—for example, he is aware of a substantial risk that the umbrella might not be his, although he thinks it probably is."); Kenneth W. Simons, *Exploring the Intricacies of the Lesser Evils Defense*, 24 LAW & PHIL. 645, 662 (2005) ("[T]o say that the actor makes a reckless mistake is to say that he is aware of a significant risk that not-P, even though, all things considered, he believes that P.").

50. *See, e.g.*, Shute, *supra* note 47, at 182–87.

51. *See, e.g.*, *id.* at 190. Or to be more precise, "[c]onsciousness is an on/off switch: a system is either conscious or not. But once conscious, the system is a rheostat: there are different degrees of consciousness." JOHN R. SEARLE, THE REDISCOVERY OF THE MIND 83 (1992).

called occurrent awareness.[52] In contrast, an actor is *dispositionally aware* of his belief that *p* if and when bringing that belief to reflective awareness would require nothing more than directing his attention to it.[53] Moreover, although not immediately present in the actor's mind, a belief with respect to which an actor is dispositionally aware is nonetheless effective or operative in guiding his conduct.

The next two states are states of *unawareness*. These states presuppose that an actor can form a belief while remaining unaware of that belief, either reflectively or dispositionally. Moreover, they presuppose that the actor is unaware of the belief because desire causes him to be unaware of it. His desire to remain unaware of the belief has managed to push it into the unconscious. An actor is *modestly unaware* of his belief that *p* if and when bringing that belief to refleetive awareness would require some effort at self-scrutiny or self-control beyond simply directing his attention to it. In contrast, an actor is *deeply unaware* of his belief that *p* if and when bringing that belief to reflective awareness would require considerable effort and outside help. A belief with respect to which an actor is deeply unaware is, one might say, a repressed belief of the sort associated in popular imagination with Freudian theory.[54]

With these additional distinctions in hand we can now say that an actor guilty of reckless homicide is one who believes, and not merely suspects, that his act is creating a risk of death, and not merely some lesser harm, where the actor is reflectively or dispositionally aware of this belief,[55] and the actor

---

52. *See* ROBERT AUDI, EPISTEMOLOGY 77 (2d ed. 2003) ("[T]o have an *occurrent property* is to be doing, undergoing, or experiencing something, as sugar undergoes the process of dissolving.").

53. *See id.* ("To have a *dispositional property* . . . is to be disposed . . . to do or undergo something under certain conditions, but not necessarily to be actually doing or undergoing or experiencing something or changing in any way."); Michael Moore, *Responsibility and the Unconscious*, 53 S. CAL. L. REV. 1563, 1621–22 (1980) ("[O]ne is conscious of going to lunch if one can state what one is doing when one's attention is turned to the subject.").

54. *See* MICHAEL BILLIG, FREUDIAN REPRESSION 14–15 (1999) (explaining that popular imagination hails Freud as the discoverer of the unconscious, where Freud insists on the distinction between the unconscious and his theory of repression).

55. Allowing dispositional awareness to suffice for reckless-homicide liability is a matter of some controversy. *See, e.g.*, Shute, *supra* note 47, at 198 (noting "little agreement amongst jurists . . . as to whether knowledge or belief may be latent, or tacit, or must be explicit for criminal liability to be established."). Some writers maintain or appear to maintain that reflective awareness, and not merely dispositional awareness, is or should be required. *See, e.g.*, DOUGLAS N. HUSAK, PHILOSOPHY OF CRIMINAL LAW 135–36 (1987) (claiming that dispositional awareness does not suffice for recklessness); Kimberly Kessler Ferzan, *Opaque Recklessness*, 91 J. CRIM. L. & CRIMINOLOGY 597, 630, 642 (2001) (arguing that the Model Penal Code's reference to "*conscious* disregard" in its definition of recklessness means reflective awareness is required); Simons, *Bad Character*, *supra* note 46, at 254 n.76 (adopting the view that the Model Penal Code definition of recklessness requires reflective awareness); Treiman, *supra* note 37, at 354 ("Having knowledge of a risk involved in specific conduct, but not being aware of the risk at the moment, that is, not stopping to think of the consequences, may be negligent but it is not reckless."); M. Wasik & M.P. Thompson, *"Turning a Blind Eye" as Constituting* Mens Rea, 32 N. IRELAND LEGAL Q. 328, 342 (1981) (endorsing requirement of reflective awareness for recklessness).

nonetheless chooses to take the risk because he seeks to satisfy a desire the law declares unworthy of satisfaction given the risk the actor believes he must take in order to satisfy it.[56] At the time he chooses to take the risk, the actor might hope the risk will not materialize, or he might be indifferent as to whether or not it does. Either way, he is still guilty of (at least) reckless homicide.[57]

Other writers maintain or appear to maintain, consistent with the position adopted here, that dispositional awareness suffices or should suffice. *See, e.g.*, R.A. DUFF, INTENTION, AGENCY & CRIMINAL LIABILITY 160–61, 164–66 (1990) (arguing that "latent knowledge" [i.e., dispositional awareness] suffices for recklessness, provided the reason the actor gave no thought to the matter was due to indifference); VICTOR TADROS, CRIMINAL RESPONSIBILITY 257 (2005) ("If conscious awareness [i.e., reflective awareness] . . . is required . . . for the proper imposition of criminal responsibility it is much too restrictive."); Alexander, *Insufficient Concern*, *supra* note 29, at 954 n.62 ("I am inclined to deem . . . low-level consciousness of risk [i.e., dispositional awareness] to be suffieient for recklessness culpability . . . ."); James B. Brady, *Recklessness*, 15 LAW & PHIL. 183, 187 (1996) (It "would be too restrictive an account of recklessness" to limit "conscious disregard" to "those cases where a person has thought about a risk."); Eric Colvin, *Recklessness and Criminal Negligence*, 30 U. TORONTO L.J. 345, 362 (1982) ("[T]here should be no question that simple consciousness of risk [i.e., dispositional awareness] falls within [the ambit of recklessness]."); Jeremy Horder, *Gross Negligence and Criminal Culpability*, 47 U. TORONTO L.J. 495, 510–11 (1997) (suggesting that "latent knowledge" [i.e., dispositional awareness] could suffice for criminal liability); Moore, *supra* note 53, at 1621 ("A principle requiring that the actor *think of* [i.e., be reflectively aware of] his actions, intentions, or reasons as he acts would be far too narrow."); Simons, *Culpability and Retributive Theory*, *supra* note 46, at 385 ("Unless a cognitive test of recklessness is somewhat expansive in its definition of . . . 'awareness,' it will have very few applications."); *cf.* Zimmerman, *supra* note 47, at 422 ("[I]n cases *other than those of routine or habitual actions*, one incurs culpability for something only if one's cognitive connection to that thing involves one's adverting to it." (emphasis added)).

56. The wrongfulness of any particular case of reckless homicide will be a function of (1) the magnitude or size of the risk of causing death (ranging anywhere from nonzero to just short of "practical certainty") that the actor believed he was taking or imposing; and (2) the content of the desire that the actor sought to satisfy in taking that risk, which desire he endorses and executes in acting. The greater the risk, the more wrongful the act (all else being equal); likewise, the more reprehensible the desire, the more wrongful the act (all else being equal).

An especially wrongful case of reckless homicide would therefore be one in which the actor believes that the probability of death is very high (though not practically certain) and the desire he seeks to satisfy in taking that risk is one the law deems especially reprehensible. Indeed, an otherwise reckless homicide of this sort would likely be graded as a form of murder, known as "depraved heart" murder at eommon law, *see* JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW § 31.05[A][1], at 512 (3d ed. 2001), "extreme indifference" murder under the MPC, *see* MODEL PENAL CODE § 210.2(1)(b) (1962), or "wicked recklessness" in Scots law. 2 SIR GERALD H. GORDON, THE CRIMINAL LAW OF SCOTLAND § 23.19, at 297 (3d ed. 2001). Liability for murder might also be warranted in cases where the desire the actor seeks to satisfy is especially reprehensible even though the risk he believes he is imposing is very small. For example, playing a one-off game of Russian roulette in which the risk of death is one-in-ten can still constitute murder inasmuch as the actor's motive for imposing that risk is nothing more than the desire to experience the thrill of endangering another's life. *Cf.* Alexander, *Insufficient Concern*, *supra* note 29, at 934 ("Imposing Russian roulette on others, no matter how high the ratio of empty to loaded chambers, seems a elear case of criminally culpable conduct."); Simons, *Bad Character*, *supra* note 46, at 263 ("[T]he Russian roulette player . . . who willingly endangers another's life for the sake of a personal thrill is extremely indifferent to the victim's fate.").

57. Some scholars argue that the wrongfulness of any particular case of reckless homicide depends not only on the magnitude or size of the risk of causing death the actor believed he was taking, nor only on the content of the desire the actor sought to satisfy in taking that risk, but also on

With this analysis of reckless homicide in place, the difference between an actor guilty of involuntary manslaughter and one guilty of reckless homicide comes into focus. The reckless actor believes that *p* and is at least dispositionally aware of his belief that *p* at the time he acts so as to create a risk of causing death. The negligent actor's cognitive state differs in one of three ways.[58]

First, he might never even form the belief that *p*, nor even the suspicion that *p*, despite the fact that his background beliefs are sufficient to support the formation of that belief.[59] Second, the actor might suspect that *p*, but never form the belief that *p*, even though his background beliefs are once again sufficient to support the formation of that belief. Third, the actor might

---

the "attitude" the actor possessed with respect to the harm risked at the time he acts. *See, e.g.*, Michaels, *supra* note 48, at 962–63; Simons, *Bad Character*, *supra* note 46, at 267–68; Kenneth W. Simons, *Rethinking Mental States*, 72 B.U. L. REV. 463, 486–90 (1992) [hereinafter Simons, *Mental States*].

According to this line of thought, an actor who takes a risk of causing death in order to satisfy a particular desire might nonetheless hope that the risk does not materialize because, though he cares too little for others to stop himself from taking the risk he needs to take in order to satisfy the desire he seeks to satisfy, he cares enough to hope the risk does not materialize. Alternatively, he might be indifferent, not caring one way or the other whether the risk materializes, because he does not care enough about others to care what happens, in which case we might say the actor accepts the risk, or is reconciled to it. For those who believe that an actor's deserved punishment depends in part on his attitude toward the harm risked, an actor who possesses an attitude of indifference toward the risk, who does not care one way or the other whether the risk materializes, deserves more punishment than does one who hopes it does not.

Yet insofar as the "attitude" an actor possesses with respect to the harm risked at the time he chooses to take that risk is ultimately reducible to desires he possesses or lacks when he acts, the punishment he deserves cannot legitimately depend on any such attitude. The desires an actor possesses or lacks at any particular point in time may show him to be a better or worse person, but unless and until his will endorses and executes a desire in action, a state committed to the principle that citizens should be punished for what they do, and not for who they are, cannot legitimately enhance an actor's punishment based simply on desires he possessed or lacked at the time he chose to take the risk. The desires an actor possesses or lacks at the time he acts may reveal the quality of his character, but they do not reveal the quality of his will. Nonetheless, if an actor's good character is a legitimate reason for extending mercy, and his bad character a legitimate reason for withholding it, an actor who hopes the risk he is imposing does not materialize might be punished less than one who is indifferent to it, not because the first actor deserves less punishment than the second, but rather because he is a more appealing candidate for mercy.

58. The analysis here is limited to cases involving ignorance (i.e., cases in which an actor fails to form the belief that *p* or forms that belief but remains unaware of it). It leaves aside any complications that might arise when the case involves an actor who has made a mistake (i.e., cases in which an actor mistakenly believes not-*p*).

59. *See, e.g.*, Victor Tadros, *Recklessness and the Duty to Take Care*, *in* CRIMINAL LAW THEORY, *supra* note 47, at 227, 257 ("If the defendant builds on [a] false belief to form the belief that there was no risk in the activity performed in this case, for the most part at least, she ought not to be regarded as [objectively] reckless [i.e., negligent]."); Low, *supra* note 19, at 549 ("Negligence . . . involves a subjective inquiry (*what the actor actually knew about the context*) and an objective inquiry (the inferences that should have drawn from what the actor knew)." (emphasis added)). *But see* A.P. Simester, *Can Negligence Be Culpable?*, *in* OXFORD ESSAYS IN JURISPRUDENCE 85, 96 (Jeremy Horder ed., 2000) ("[T]he pool of information for evaluating action is the *union* of that information which she *should* reasonably *have possessed*, and that which she did in fact possess." (second emphasis added)).

form the belief that *p*, but fail to become aware of that belief. As we will see, actors possessing the second and third of these cognitive states, neither of which suffices for reckless homicide liability, are nonetheless fair candidates for involuntary manslaughter liability while actors in the first category are not.

Although the cognitive states of the negligent actor and the reckless actor differ in one of the ways just described, they are in one important respect the same. The background beliefs available to both actors are sufficient to support the formation of the belief that *p* at the moment each acts so as to unleash a risk of causing death. Indeed, the reckless actor does form that belief. But what if an actor's background beliefs are themselves not what they should be? What if the actor possesses background beliefs he should not, or does not possess background beliefs he should? What then? We can sort the cases into two categories.

The first covers *exculpatory* cases. In these cases the actor fails to form the belief that *p*, not because of any defect in the belief-formation process itself, but rather because the actor possesses a background belief he should not possess (mistake) or does not possess a background belief he should possess (ignorance). Consequently, the beliefs available to the actor are insufficient to support the formation of the belief that *p*. In fact, they might instead support the formation of the belief that not-*p*. In other words, given the beliefs available to him at the time, the actor could not have formed the belief that *p*. Thus, insofar as ought implies can, his ignorance with respect to *p* at the moment he should have formed the belief that *p* is excusable ignorance.

Accordingly, such an actor should not be held liable for involuntary manslaughter, let alone reckless homicide. Involuntary manslaughter liability should be limited to those actors whose available background beliefs *do* support the formation of the belief that *p* but who culpably fail to form that belief or who form that belief but culpably remain unaware of it. The culpability, if any, of an actor for forming background beliefs he should not have formed or failing to form background beliefs he should have formed—the presence or absence of which in turn preclude the formation of the belief that *p*—is a failing separate and distinct from his failure to form the belief that *p* based on the background beliefs available to him at the time. The culpability associated with any such prior failing should be assessed on its merit.[60] The same is true of background beliefs the actor forms but then forgets. Once forgotten those beliefs are effectively unavailable to support the formation of the belief that *p*.

60. An actor may be independently culpable for having done something to cause his background ignorance or mistake, or for having failed to do something to remedy his background ignorance or mistake, but he is not culpable for the causal effect of his ignorance or mistake on his subsequent failure to form the belief that *p*. *See infra* notes 69–82 and accompanying text (discussing the prior-choice theory).

The second covers *inculpatory* cases. In these cases the actor *does* form the belief that *p*. Yet he forms that belief, not because of any defect in the belief-formation process itself, but rather because he possesses a background belief he should not possess (mistake), or does not possess a background belief he should possess (ignorance). Only now his mistake or ignorance is inculpatory, not exculpatory. The actor believes that his act is creating a risk of causing death, although he would not so have believed if he had possessed the background beliefs he should have possessed or did not possess those he should not have possessed. Nonetheless, he *does* believe that *p*. Consequently, if he chooses to take the risk he believes exists in order to satisfy a desire the law condemns, then he should be held liable for reckless endangerment or attempted reckless homicide if death does not result,[61] or for reckless homicide if it does.[62]

Cases in which criminal liability for involuntary manslaughter is warranted fall into neither of these categories. Involuntary-manslaughter liability is warranted only where the background beliefs available to the actor support the formation of the conscious belief that *p* (unlike the exculpatory case), but the actor fails to form that belief (unlike the inculpatory case). The culpability of an actor who unwittingly creates an unjustified risk of causing death consists not in the possession of background beliefs he should not have possessed, nor in the failure to possess background beliefs he should have possessed. It consists instead in his failure to form the conscious belief that *p* when the background beliefs available to him support its formation, and when he could and fairly should have formed it based on those background beliefs.

---

61. *See* Robinson, *supra* note 39, at 387 (arguing that such an actor would be guilty of reckless endangerment under the Model Penal Code). The Code defines "Recklessly Endangering Another Person" as "recklessly [engaging] in conduct which places or may place another person in danger of death or serious bodily injury." MODEL PENAL CODE § 211.2 (1962). I should note that most jurisdictions do not recognize attempted reckless homicide as a crime. *See* Michael T. Cahill, *Attempt, Reckless Homicide, and The Structure of Criminal Law*, 78 U. COLO. L. REV. (forthcoming 2007) (manuscript at 3), *available at* http://ssrn.com/abstract=894379 ("In nearly all jurisdictions to consider the question, courts have held that no such offense [of attempted reckless homicide] exists.").

62. *See* Robinson, *supra* note 39, at 387 (arguing that such an actor would be guilty of reckless homicide (manslaughter) under the Model Penal Code); *see also* Alexander, *Insufficient Concern*, *supra* note 29, at 936 (arguing that such an actor should be guilty of reckless homicide). The longstanding and ongoing debate between those who argue that the extent of an actor's punishment should reflect the harm he actually and proximately causes and those who argue that it should reflect only the harm he intends to cause or consciously risks causing can safely be left to one side for present purposes. *Compare* MICHAEL MOORE, PLACING BLAME 193 (1997) ("[C]ulpability is both independently necessary and independently sufficient as a basis for some punishment, whereas wrongdoing [or harm-causing] is neither. Yet wrongdoing is an independent desert basis . . . : when culpability is present, wrongdoing independently influences how much punishment is deserved."), *with* Stephen J. Morse, *Reason, Results, and Criminal Responsibility*, 2004 U. ILL. L. REV. 363, 428 (2004) (advocating an approach according to which "[t]here would be no difference in blame and punishment between conduct that causes harm and similar conduct that does not"). My current sympathies rest with the position Morse defends.

II. Retribution and Involuntary Manslaughter

If an actor unleashes into the world a risk of causing another's death (and death results), and if the background beliefs available to him at the time support the formation of the belief that *p*, but the actor either fails to form that belief or forms that belief but fails to become aware of it, should he be punished? If so, why?

For those who embrace utilitarianism, he probably should be punished. On this view the state is obligated to punish an actor when and because the good consequences of punishing him outweigh the bad.[63] If punishing the unwitting actor happens to produce such a result, utilitarianism has no trouble explaining why he should be punished. One might argue, for example, that his punishment will create the incentive needed to encourage him and others similarly situated to exercise due care in the future. That may be, but utilitarianism cannot avoid the charge that it countenances the punishment of the innocent and the disproportionate punishment of the guilty.[64] It would punish the unwitting actor even though he does not *ex hypothesi* deserve to be punished, which is reason enough to set utilitarianism aside.

The natural alternative to utilitarianism is retribution. On this view the state is obligated, all else being equal, to punish an actor because and to the extent, but only to the extent, that he deserves to be punished.[65] Retributive theories differ from one another in (among other ways) what they take to be the basis of one's just deserts. The orthodoxy draws a distinction between character retribution and choice retribution.

Character retribution says that deserved punishment is ultimately based on the odious character an actor reveals when he commits a crime without justification or excuse. Insofar as an actor does indeed display sueh a character when he fails to realize that he is imposing on others an unjustified risk of causing death—he is careless, thoughtless, indifferent, and so forth—character retribution likewise has no trouble explaining why such an actor

63. *See, e.g.*, MOORE, *supra* note 62, at 91 ("[T]he ultimate justification for inflicting the harm of punishment is that it is outweighed by the good to be achieved.").

64. *See, e.g.*, *id.* at 94–97. So-called negative retributive theories according to which the pursuit of utilitarian ends is subject to limits or side constraints barring the punishment of the innocent and the disproportionate punishment of the guilty are an improvement on pure utilitarian theories. They are nonetheless vulnerable to other objections. *See id.* at 97–102. For a recent effort to defend the claim that one utilitarian goal (specific deterrence) should be the criminal law's "predominant goal," see Christopher Slobogin, *The Civilization of the Criminal Law*, 58 VAND. L. REV. 121 *passim* (2005).

65. *See, e.g.*, MOORE, *supra* note 62, at 104 ("Retributivism . . . is the view that it is a sufficient reason for us to have punishment institutions (i.e., the criminal law) . . . that the person deserve to be punished."). It would be more precise to say that the statement in the text describes positive retribution, according to which the state is obligated to impose retributive punishment, in contrast to negative retribution, according to which the state is permitted though not obligated to impose retributive punishment.

should be punished.[66] He deserves to be punished because he has shown himself to be possessed of a character worthy of condemnation. That may be, but character retribution is among other things inconsistent with the general principle, to which any liberal state owes allegiance, that citizens can legitimately be punished only for what they do, not for who they are.[67]

That leaves choice retribution, according to which deserved punishment is ultimately based on an actor's choice to do that which the law demands he not do. But choice retribution and punishment for unwitting risk imposition don't mix, since an actor who does not realize he is imposing a risk of causing harm cannot intelligibly be said to have chosen to take that risk. Consequently, choice retribution would rule out punishment for involuntary manslaughter altogether. Indeed, many scholars who can be understood as proponents of choice retribution have reached precisely that conclusion.[68]

---

66. Character retribution will of course have trouble explaining why an actor who unwittingly imposes a lethal risk should be punished when the actor's failure to realize the risk was due to a culpable but momentary lapse not sufficiently stable or enduring to be fairly characterized as a trait of character at all. The fact that an actor may display indifference on a particular occasion does not necessarily mean he is characteristically indifferent. *See, e.g.*, MOORE, *supra* note 62, at 590 ("[J]udgements of negligence do not depend on any general traits of carelessness; an isolated act suffices for responsibility for the harm that act causes."); Simons, *Culpability and Retributive Theory*, *supra* note 46, at 389 ("[I]ndifference can be illustrated even in a single 'out of character' act.").

67. *See* MOORE, *supra* note 62, at 586 ("[N]o onc deserves to be punished for being a poor specimen of humanity. The aesthetic kind of responsibility that we admittedly do have for so much of our characters as are unchosen cannot fairly lead to punishment, because we could not have avoided possessing these aspects of ourselves."); Jeffrie G. Murphy, *The State's Interest in Retribution*, 5 J. CONTEMP. LEGAL ISSUES 283, 298 (1994) (concluding that robust versions of retributivism, including character retributivism are probably not compatible with liberalism).

68. In contemporary American scholarship, this position is prominently associated with the work of Larry Alexander. *See* Alexander, *Reconsidering the Relationship*, *supra* note 29, at 85 (concluding that the voluntary act principle requires us to reject principles that envision crimes predicated on negligence); *see also* Larry Alexander, *Crime and Culpability*, 4 J. CONTEMP. LEGAL ISSUES 1, 7 (1994) [hereinafter Alexander, *Crime and Culpability*] ("[A]lthough some cases of inadvertent negligence reflect prior culpable acts . . . inadvertent negligence is not itself culpable."); Alexander, *Insufficient Concern*, *supra* note 29, at 952 ("Because the actor who fails to advert to a risk acts in ignorance of that risk, he is not culpable for taking it.").

Earlier scholarship also defended this position. *See* J.W.C. Turner, *The Mental Element in Crimes at Common Law*, 6 CAMBRIDGE L.J. 31, 39 (1936) ("It is . . . submitted with emphasis that although . . . negligence may be blameworthy and may ground civil liability, *it is at the present day not sufficient to amount to* mens rea *in crimes at common law*."); *see also* Jerome Hall, *Negligent Behavior Should Be Excluded from Penal Liability*, 63 COLUM. L. REV. 632, 632 (1963) ("[I]t would be a great step forward to exclude negligent behavior from the scope of penal liability."). H.L.A. Hart, in an influential reply to Turner, argued that Turner's arguments "rest on a mistaken conception both of the way in which mental or 'subjective' elements are involved in human action, and of the reasons why we attach the great importance which we do to the principle that liability to criminal punishment should be conditional on the presence of a mental element." H.L.A. HART, *Negligence,* Mens Rea, *and Criminal Responsibility*, *in* PUNISHMENT AND RESPONSIBILITY 136, 139 (1968). See Finkelstein, *supra* note 46, at 581, for a contemporary defense of Turner against Hart, in which she argues that "[l]iab[ility] in the absence of . . . subjective awareness on [an actor's] part of what he was doing, under its prohibited description[,] . . . is objectionable because it is liability in the absence of ordinary responsibility." James Brady and A.P. Simester have replied to Hall and Finkelstein. *See* James B. Brady, *Punishment for Negligence: A Reply to Professor*

For them, any punishment of an actor who unwittingly creates a risk of causing another's death is undeserved and hence unjust, because the actor did not choose to take that risk.

One might nonetheless try to reconcile choice retribution with criminal liability for unwitting risk-creation in either of two ways. The prior-choice theory looks to the actual past. It asks: Did the actor make some prior culpable choice to impose on others an unjustified risk of causing death, or at least some prior culpable choice but for which he would have realized he was imposing such a risk? If so, then punishment is warranted. The hypothetical-choice theory looks to an imagined past. It asks: Would the actor have chosen to take the risk if he had been aware of it, even though he was not in fact aware of it? If so, then again, punishment is warranted. In the end, however, neither theory accurately identifies when or why the state may legitimately punish an actor who unwittingly creates a lethal risk.

A. The Prior-Choice Theory

The prior-choice theory presupposes that an actor who unwittingly creates a lethal risk cannot fairly be subject to criminal liability for imposing such a risk because, being unaware of the risk, he did not choose to impose it. The theory likewise presupposes that the actor's failure to become aware of the risk (i.e., to form the conscious belief that p) is not itself culpable, because at the time the actor should have formed the belief that p, he lacked the capacity to form that belief, and lack of capacity is ordinarily an excuse. Ought implies can, at least when the reaction to an actor's failure to do as he ought is state punishment.

An actor can lack the capacity to form the belief that p at time t_1 for one of three reasons. First, the actor can lack the capacity to form the belief that p because the background beliefs available to him were insufficient to support the formation of the belief. Second, assuming that the background beliefs internally available to the actor were sufficient to support the formation of the belief that p, the actor can nonetheless lack the capacity to form the belief that p because he lacked the cognitive capacity needed for the belief to form. He lacked the requisite intellectual wherewithal. Third, assuming that the background beliefs available to the actor were sufficient to support the formation of the belief that p, and that the actor possessed the cognitive capacity needed to form that belief based on those background beliefs, the actor can nonetheless lack the capacity to form the belief that p because he lacked the conative capacity needed for the belief to form. He did

---

Hall, 22 BUFF. L. REV. 107, 121 (1972) (arguing that the "failure to exercise normal capacities to avoid harm makes [a negligent actor's] conduct blameworthy," Hall's claims to the contrary notwithstanding); A.P. Simester, *Responsibility for Inadvertent Acts*, 2 OHIO ST. J. CRIM. L. 601, 606 (2005) ("[N]ot all inadvertent acts are culpable—we acknowledge this, in legal usage, with the familiar distinction between negligence and accident. One does not need to think of everything. But on Finkelstein's account, we need think of nothing.").

not care enough, either for the truth or for the well-being of others, needed for him to have realized the risk he was creating.

Because the prior-choice theory assumes that the actor at time t_1 had a valid excuse (based on lack of capacity) for not forming the belief that p, the theory looks back in time for some prior culpable choice with respect to which the actor had no excuse.[69] From here the prior-choice theory splits into two. The narrow version looks for a prior culpable choice on the actor's part to violate the obligation not to act so as to create an unjustified risk of causing death. The broad version looks for a prior culpable choice to violate a range of other obligations, where that violation in turn causes the actor's later incapacity to form the belief that p at time t_1.

1. The Narrow Prior-Choice Theory.—According to the narrow prior-choice theory, an actor who unwittingly creates an unjustified risk of causing death (with death resulting) cannot fairly be punished if and because he lacked the capacity at time t_1 to form the belief that he was creating a lethal risk, and thus cannot be said to have chosen to take that risk. Nonetheless, he *can* fairly be punished if he chose to impose a lethal risk at some earlier point

---

69. *See, e.g.*, Alexander, *Crime and Culpability*, *supra* note 68, at 7 ("[A]lthough some cases of inadvertent negligence reflect prior culpable acts . . . , inadvertent negligence is not itself culpable."); Alexander, *Insufficient Concern*, *supra* note 29, at 950 ("Of course, inadvertence to risk may reflect moral culpability in the sense that it is the product of a prior reckless act."); Alexander, *Reconsidering the Relationship*, *supra* note 29, at 101–04 ("[T]here may be a sizeable subclass of negligent acts where the failure of the defendant to advert to the risk was caused by some prior culpable—and hence fully voluntary—choice of the defendant."). Statements of the prior-choice theory also appear in the philosophical literature on moral responsibility. *See* Ishtiyaque Haji, *An Epistemic Dimension of Blameworthiness*, 57 PHIL. & PHENOMENOLOGICAL RES. 523, 541–42 (1997) (arguing that an actor who does not believe an act is wrong is not blameworthy for performing that action, though he may be blameworthy for performing earlier actions he does believe are wrong and that result in his not believing the later action is wrong); Gideon Rosen, *Culpability and Ignorance*, 103 PROC. ARISTOTELIAN SOC'Y 61, 63 (2002) ("If I recklessly shirk [an epistemic] obligation and wind up ignorant as a result, the ignorance itself is culpable, and in that case it's no excuse." (emphasis omitted)); Holly Smith, *Culpable Ignorance*, 92 PHIL. REV. 543, 570 (1983) ("In cases of culpable ignorance, the [later] unwitting act is a risked upshot of the [prior] benighting act."); Zimmerman, *supra* note 47, at 417 ("[C]ulpability for ignorant behavior must be rooted in culpability that involves no ignorance."). See James A. Montmarquet, *Culpable Ignorance and Excuses*, 80 PHIL. STUD. 41 (1995) [hereinafter Montmarquet, *Culpable Ignorance*] and James Montmarquet, *Zimmerman on Culpable Ignorance*, 109 ETHICS 842 (1999), for critiques of Smith and Zimmerman, respectively. For Zimmerman's response to Montmarquet, see Michael J. Zimmerman, *Controlling Ignorance: A Bitter Truth*, 33 J. SOC. PHIL. 483, 485 (2002).

Steven Sverdlik argues that an actor's failure to advert to a risk can be culpable even if it cannot be traced back in time to any prior culpable choice, or in other words, even if the actor's failure to advert to the risk is "pure." *See* Steven Sverdlik, *Pure Negligence*, 30 AM. PHIL. Q. 137, 141 (1993). According to Sverdlik, cases of pure negligence are culpable insofar as the actor could have engaged in the "moral reflection" needed for him to have become aware of the risk he was creating. *See id.* at 142. Although I agree with Sverdlik's claim that an actor's failure to advert to a risk can be culpable even if his ignorance cannot be traced back in time to any prior culpable choice, I disagree with his claim that an actor's failure to exercise his capacity for "moral reflection" is sufficient to render his ignorance culpable. *See infra* notes 100–12 and accompanying text (describing conditions under which an actor's failure to form the conscious belief that p is culpable).

in time (t_{1-n}). If he chose to act at time t_{1-n} with the purpose of causing death, or believing that death either would or might result, then assuming his act is neither justified nor excused, he can fairly be held liable for murder or reckless homicide if death results, or for attempted murder or attempted reckless homicide if it does not.

In other words, when viewed up close, some cases of murder or reckless homicide might mistakenly look like cases in which the actor's risk imposition should be excused for lack of capacity. The true picture emerges only when we pull back and open the time frame. When we do so we see that sometimes the actor has made a prior culpable choice rendering him liable for murder or reckless homicide, and not simply involuntary manslaughter. For example, if the Williamses had realized at time t_{1-n} that inaction would risk their son's life, they would have been guilty of reckless homicide (assuming their inaction was unjustified), even if somehow they later came to lack the belief that his life was in danger, or even came to believe that his life was in no danger.[70] Of course, if the actor has made no such prior culpable choice, and remained unaware of the risk he was creating or allowing to persist, then no liability would be warranted.

The narrow prior-choice theory is unobjectionable, at least insofar as the liability it would impose on an actor matches the culpability associated with his prior choices. The problem is that it fails to explain when and why an actor's failure to realize he is creating a risk of causing death is culpable when his ignorance cannot be traced back to any prior culpable choice. It wrongly assumes that an actor's prima facie lack of capacity to form the belief that p at time t_1 necessarily excuses his failure to form that belief, thus triggering the need to look back in time for some prior culpable choice to impose a lethal risk. It ignores the possibility that an actor may sometimes be able through the exercise of doxastic self-control to remove the impediments causing him otherwise to lack the capacity to form the conscious belief that p, such that his culpability for failing to form that belief consists in his failure to exercise such self-control, and not in any prior choice. The basic idea is that the actor can and sometimes should intervene into his own self-constitution, that doing so will enable him to see the risk he is creating, and that his failure to so intervene is culpable.

---

70. The case usually cited as an example of the narrow prior-choice theory is *People v. Decina*, 138 N.E.2d 799 (N.Y. 1956), in which Decina, an epileptic, suffered a seizure while driving his car. *Id.* at 800. The car went out of control and four children were killed. *Id.* at 801. Decina claimed that he had a valid defense to the crime charged and moved to dismiss the indictment: while in the midst of the seizure he lacked the capacity to control his car and so could not have avoided creating the lethal risk. *Id.* at 803. The Court of Appeals upheld the order of the Appellate Division upholding in turn the trial court's refusal to dismiss the indictment. *Id.* at 804. Decina knew he suffered from epilepsy, knew he might at any moment have a seizure, and nonetheless chose to get into his car and drive. *Id.* at 803–04. Although he did not choose to let his car go out of control at time t_1, he nonetheless chose to create a lethal risk at time t_{1-n} when he chose to drive, and that choice ultimately caused the children's deaths.

2. The Broad Prior-Choice Theory.—The broad prior-choice theory focuses, not on an actor's prior choice to violate the obligation against creating an unjustified risk of causing death, but on some other obligation the breach of which is the but-for and proximate cause of the actor's later incapacity at time t_1 to form the belief that p.[71] Although lack of capacity would ordinarily be an excuse, the broad prior-choice theory holds that an actor forfeits to some extent or another any such excuse if and because his incapacity at time t_1 is the result of a prior culpable choice.[72]

A complete account of the broad prior-choice theory would need to address two issues. First, it would need to say what other obligations we have, the culpable breach of which will result in forfeiture of an otherwise valid incapacity excuse. These obligations might be specified ex ante in the text of a criminal code, or a judge or jury might determine them ex post in the course of a criminal prosecution.[73] The content of these obligations might range anywhere from an obligation to gather evidence and information, so as to have available at time t_1 all the evidence one ought to have at that time, to an obligation to develop in oneself those capacities, both cognitive and conative, that one ought to possess and that one would need to possess in order to form the belief that p based on the evidence available at time t_1.[74] The relevant obligation might also be an obligation to *think* or

71. Again, an actor might lack the capacity to form the belief that p either because he then and there lacks the background beliefs sufficient to support the formation of that belief or because he lacks the cognitive or conative capacities necessary for its formation. *See supra* subpart II(A).

72. For the classic discussion of this general problem, see Paul H. Robinson, *Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine*, 71 VA. L. REV. 1 (1985).

73. *See, e.g.*, Alexander, *Reconsidering the Relationship*, *supra* note 29, at 102–03 (describing these two approaches). Allowing the existence and content of an obligation to be determined ex post naturally raises legitimate concerns about holding an actor criminally liable for breaching an obligation about which he might reasonably have been unaware at the time he acted or failed to act. Such concerns are especially acute when the obligation breached is an obligation to act, such that the breach arises from an omission. *See, e.g.*, Larry Alexander, *Criminal Liability for Omissions: An Inventory of Issues*, *in* CRIMINAL LAW THEORY, *supra* note 47, at 121, 124 ("Because the duty to rescue exists only when the rescue is 'safe' or 'easy'—and because its existence also depends on some very vague notions of status relationships and reliance—the scope of the duty is arguably so vague as to threaten its legality.").

74. Aristotle is often read to have embraced a position along these lines. *See* 2 ARISTOTLE, *Nicomachean Ethics*, *in* THE COMPLETE WORKS OF ARISTOTLE III.5.1113b–1114a, at 1729, 1758 (Bollingen Series No. 71, Jonathan Barnes ed., 1984) (Revised Oxford Translation). This position can also be found in the work of contemporary writers. *See, e.g.*, Jean Hampton, Mens Rea, SOC. PHIL. & POL'Y, Spring 1990, at 1, 27 ("If we take an Aristotelian attitude towards [an] inadvertently negligent person, we believe that, although she did not know that what she was doing was wrong, that ignorance was a function of her faulty character . . . . So we locate the defiance not at the time of the act, but earlier, during the process of character-formation."); Hilary Kornblith, *Justified Belief and Epistemically Responsible Action*, 92 PHIL. REV. 33, 38 (1983) ("[T]here can be no doubt that one can self-consciously instill in oneself a certain circumspection in circumstances where [an epistemic] mistake is likely to occur . . . ."); Simester, *supra* note 59, at 98 ("Our duty to take care . . . obligates us also to acquire the capacity to do so; in effect, we may be expected to ascertain our faults and to remedy or at least compensate for them."); Tadros, *supra* note 59, at 246 ("That we have some control over whether we have the vice of clumsiness makes [an act attributable to that

reason or *deliberate*, at least insofar as thinking, reasoning, or deliberating can be characterized as actions, on the theory that that actor would have formed the belief that *p* if he had so thought, reasoned, or deliberated.[75]

For example, perhaps the Williamses should have taken a parenting class on children's medical emergencies, where they would have been exposed to information enabling them to identify life-threatening risks.[76] If their failure to take the class and thereby gain the relevant information was a but-for and proximate cause of their later incapacity to realize the danger to their son's life, then they would forfeit the opportunity to claim this lack of capacity as an excuse to their subsequent failure to form the belief that *p*.

Second, assuming some prior culpable choice can be identified, the theory would need to say what effect, if any, that choice will have on the actor's opportunity to claim the incapacity excuse to which he would otherwise be entitled. One option would be to say that the actor's prior culpable choice should result in a partial forfeiture of his opportunity to claim lack of capacity as an excuse, such that his lack of capacity affords him only a partial excuse.[77] In this case the actor would presumably be held liable, not for reckless homicide, but for the lesser offense of involuntary manslaughter. Another option would be to say that the actor's prior culpable choice should result in the complete forfeiture of this opportunity to claim lack of capacity as an excuse, such that his lack of capacity affords him no excuse. In this case he would presumably be held liable for reckless homicide.[78]

---

vice] reflect our agency sufficiently to attribute us with responsibility for doing so."). See Kyron Huigens, *Virtue and Criminal Negligence*, 1 BUFF. CRIM. L. REV. 431, 444 (1998), for an alternative reading of the relevant passages of Aristotle's *Ethics*, according to which "actions done in ignorance . . . are . . . not subject to blame, only if . . . the agent [subsequently] feel[s] regret for having committed them."

75. *See, e.g.*, AUDI, *supra* note 52, at 160 ("Reasons, one might say, can lead to inferential beliefs by two different paths, one requiring reason*ing*, the other not. . . . In the first case, I *do* something . . . . In the second, something *happens* in me . . . ."); JOSEPH RAZ, ENGAGING REASON 13 (1999) ("We can choose to perform or not to perform mental acts, in which action has to do with thinking . . . . [But] while we can choose when to think and of what, it does not follow that we can choose the results of our thinking."). Holly Smith seems to believe that an actor's drawing of an incorrect inference, or presumably his failure to draw a correct inference, can also qualify as a culpable act or omission. *See* Smith, *supra* note 69, at 547 (arguing that a prior culpable act, which she refers to as a "benighting act" can be a "mental occurrence (such as making an incorrect inference), and the temporal gap between it and the unwitting act is infinitesimal"). I am presently inclined to think that inference-drawing or failing to draw an inference cannot be fairly characterized as an act or omission insofar as neither is subject to the will. Thinking, deliberating, and so forth are mental acts subject to the will, but belief-formation itself is not. Thus, whether one characterizes inference-drawing as an act or not may depend on whether one uses the phrase "inference-drawing" to mean thinking or deliberating, or whether one uses the phrase to refer to belief-formation itself.

76. Alexander, *Reconsidering the Relationship*, *supra* note 29, at 101–02.

77. *See, e.g.*, Smith, *supra* note 69, at 549 (describing this view as the "Moderate View").

78. *See, e.g.*, *id.* at 548 (characterizing this view as the "Conservative View"). An actor would remain free to claim incapacity as an excuse if (1) he has made no prior culpable choice upon which a forfeiture can be grounded; or (2) he has made a prior culpable choice but that choice is given no effect with regard to the actor's opportunity to assert the incapacity excuse to which he is otherwise

The Model Penal Code rule governing voluntary intoxication provides one example of the broad prior-choice theory in practice.[79] The voluntarily-intoxicated actor chooses to become or risk becoming intoxicated.[80] If he later acts so as unwittingly to create an unjustified risk of causing death (and death results), he would, but for the operation of the Code's forfeiture rule, have a defense to a charge of reckless homicide. He lacked the awareness of the risk required to describe him as reckless. But the Model Penal Code treats his prior choice to risk intoxication as the basis upon which he forfeits that defense inasmuch as his failure to be aware of the risk is, if due to his intoxication, "immaterial."[81] Even if he lacked the capacity to form the conscious belief that p at the time he created the risk, he forfeits the opportunity to substantiate that defense based on his prior choice to risk intoxication.

Unlike the narrow prior-choice theory, the broad prior-choice theory *is* objectionable. The problem lies in the forfeiture rules on which the theory is based. Such rules result in an actor's being punished even though he could not have realized at time t_1 that he was creating a risk of causing death and *a fortiori* never chose to impose any such risk. An actor who falls within the scope of a forfeiture rule is therefore one who would otherwise have a valid defense to liability. He is nonetheless partially or completely divested of that defense due to some prior culpable choice. Consequently, assuming the culpability associated with the actor's prior choice is less than the culpability associated with the crime of which he is convicted, punishing him for that crime would amount to disproportionate and thus unjust punishment, though how disproportionate and unjust will depend on the culpability of the prior choice.[82] A parent should not go to prison for involuntary manslaughter because he or she decided to skip class one day.

---

entitled, although the actor remains liable for his prior culpable choice itself. *See, e.g., id.* at 549 (describing (2) as the "Liberal View").

79. *See* MODEL PENAL CODE § 2.08 (1962). The exact rules governing the effect of voluntary intoxication on an actor's criminal liability vary from jurisdiction to jurisdiction. *See, e.g.*, Mitchell Keiter, *Just Say No Excuse: The Rise and Fall of the Intoxication Defense*, 87 J. CRIM. L. & CRIMINOLOGY 482, 518–20 (1997) (categorizing various state-law rules governing the admissibility of evidence of voluntary intoxication in criminal prosecutions).

80. The Code actually treats an actor as suffering from "self-induced intoxication" when he "knowingly introduces [substances] into his body, the tendency of which to cause intoxication he . . . *ought to know*." MODEL PENAL CODE § 2.08(5)(b) (1962) (emphasis added). The Code therefore treats negligence with respect to the intoxicating nature of a knowingly-introduced substance as a sufficient basis for the operation of the forfeiture stated in § 2.08(2). Recklessness with respect to the intoxicating nature of the substance would also suffice. *See id.* § 2.02(5).

81. *See id.* § 2.08(2).

82. *See* Robinson, *supra* note 72, at 29 ("To punish an actor for conduct performed when he is not responsible is to punish him for conduct that society has determined to be . . . blameless."). Involuntary-manslaughter liability might also be portrayed as resting on a so-called substitution rule, according to which one mental state or set of mental states is allowed to substitute for another. As Michael Moore explains it:

> On this view negligence is just a substitution rule for belief states: when an actor believes that he is shooting at a target, believes he hears a shout of 'don't shoot', believes only humans make such sounds and usually only when in danger, yet does not

In any event, like the narrow prior-choice theory, the broad prior-choice theory cannot explain when and why an actor's failure to realize he is creating a risk of causing death is culpable when his ignorance cannot be traced back to any prior culpable choice. It wrongly assumes, as does the narrow prior-choice theory, that an actor's prima facie lack of capacity to form the belief that p at time t_1 necessarily precludes any culpability for the actor's failure to form that belief. The broad prior-choice theory must therefore impute culpability for the actor's lethal risk-creation where no such culpability exists. It ignores, as does the narrow prior-choice theory, the possibility that an actor may sometimes be able through the exercise of doxastic self-control to remove the impediments causing him to lack the capacity to form the conscious belief that p, such that his culpability for failing to form that belief consists in his failure to exercise such self-control, and not in any prior choice for which he is then held to forfeit the opportunity to plead or prove lack of capacity as a defense.

Can the prior-choice theory explain why one might think retributive punishment is warranted for Sam and Tiffany but not for the Williamses? It depends. If Sam and Tiffany are guilty of a prior culpable choice that in turn caused them to be unaware of the risk to their child's life, but the Williamses are not guilty of any such prior choice, then the theory can explain the difference. But insofar as the difference between the two cases is thought not to depend on any such contingency, then the explanation the prior-choice theory offers for the difference is the wrong explanation.

B. The Hypothetical-Choice Theory

Another way in which one might try to reconcile choice retribution with criminal liability for unwitting risk creation ultimately turns not on actual prior choice, but on imagined hypothetical choice.[83] The hypothetical-choice

---

> believe that his shooting risks becoming a killing, he is culpable for believing (to the probability of reasonable inference) that he is killing.

MOORE, *supra* note 62, at 413. In other words, an actor whose background beliefs support the formation of the belief that p will be treated as if he believed that p, even though he did not in fact believe that p. The background beliefs he has substitute for the belief that p, which he should have had but did not. The problem with such a rule, as with the forfeiture rules discussed in the text, is that it results in punishment disproportionate to culpability because the actor is treated as if he possessed a culpable mental state when he in fact did not. *See id.* at 413 ("One who intends to inflict grievous bodily harm has a level of eulpability commensurate to that wrong, not a level of culpability commensurate to the wrong actually done but not intended."); *see also* Ferzan, *supra* note 55, at 603–06 (arguing that substituting one mental state for another "demonstrates a willingness to punish people disproportionate to their culpability").

83. The description of the hypothetical-choice theory offered in the text is based on arguments contained in the writings of a number of scholars. For reasons that will become apparent this theory might also be called the "indifference theory." *See, e.g.*, MAYO MORAN, RETHINKING THE REASONABLE PERSON 257 (2003) (describing this theory as the "indifference account").

The hypothetical-choice theory arguably once had and arguably continues to have some support in English law. In *Regina v. Caldwell*, [1982] A.C. 341 (H.L. 1981) (appeal taken from Eng.), the House of Lords held that an actor could be said to be reckless as to some harm if "(1) he does an act

theory posits that some but not all cases in which an actor unwittingly creates an unjustified risk of causing death warrant retributive punishment. What distinguishes those cases in which punishment is warranted from those in which it is not is *why* the actor failed to form the belief that *p*.[84] When the beliefs available to an actor support the formation of the belief that *p*, an actor's failure to form that belief can have one of two explanations.

First, the actor may lack the cognitive capacity needed to form the belief that *p*.[85] Whether an actor lacks the necessary cognitive wherewithal to form the belief that *p* will sometimes be hard to say, but if he does, then he could not form the belief that *p*, which explains why he did not form it. Second, the actor may lack the conative capacity needed to form the belief that *p*. The idea here is that some standing desire to treat and regard others as moral agents worthy of concern and respect is needed in order for an actor to realize that he is creating a risk of causing harm to others. An actor suffers from a conative incapacity to form the belief that *p* if, at the moment he should have formed that belief, his conative makeup prevents its formation, even though his background beliefs support its formation and even though he possesses the cognitive capacity to form it.

The hypothetical-choice theory is based on the second explanation. It presupposes that an actor who fails to realize he is creating a lethal risk possesses the cognitive capacity to form the belief that *p*, but nonetheless lacks the requisite conative capacity.[86] He fails to realize he is creating a lethal risk because, one might say, he is *indifferent* to the well-being of others, or

---

which in fact creates an obvious risk [of causing that harm] and (2) when he does the act *he* either *has not given any thought to the possibility of there being any such risk* or has recognised that there was some risk involved and has none the less gone on to do it." *Id.* at 354 (opinion of Lord Diplock) (emphasis added). *Caldwell* provoked considerable controversy among English commentators. *See* Simons, *Mental States*, *supra* note 57, at 487–88 n.87 (collecting literature). Although *Caldwell* now only applies to a very few statutory offenses in England, *see* ASHWORTH, *supra* note 38, § 5.3(c), at 187 ("The *Caldwell* definition is now of little practical significance."), A.P. SIMESTER & G.R. SULLIVAN, CRIMINAL LAW § 5.2(i), at 143 (2d ed. 2003) (noting that in England the *Caldwell* decision is applicable to few offenses other than the offense in *Caldwell* itself and that *Caldwell* no longer applies to the offenses of manslaughter, assault, or rape), the arguments of hypothetical-choice theorists, especially English theorists, can be understood as efforts to vindicate *Caldwell*'s basic proposition that some cases of unwitting risk imposition warrant retributive punishment. *See, e.g.*, DUFF, *supra* note 55, at 165 ("The truth in *Caldwell* . . . is that the indifference which constitutes recklessness can sometimes be shown in an agent's very failure to notice a risk which her action creates, as well as in her conscious risk-taking.").

84. DUFF, *supra* note 55, at 165–66 ("[W]hat matters is not just *that*, but *why*, the agent fails to notice an obvious risk; she is reckless [culpably negligent] only if she fails to notice it *because* she does not care about it.").

85. *See, e.g.*, Simester, *supra* note 59, at 102–03 (distinguishing low intelligence from inattention to demonstrate the difference in underlying beliefs between the mistake understandably made by an actor of low intelligence and a mistake carelessly made by an inattentive actor of higher intelligence).

86. *See, e.g.*, DUFF, *supra* note 55, at 154 (describing "indifference" in terms of "attitudes" or "feelings"); Horder, *supra* note 55, at 501 (describing "indifference" as an "affective state"); Simons, *Culpability and Retributive Theory*, *supra* note 46, at 388 (describing "indifference" as a "desire-state").

lacks sufficient concern for their well-being, or *does not care enough* about their well-being.[87] The conative constitution of such an actor can be configured in two different ways, either of which results in his lacking sufficient concern for the well-being of others at time t_1 and thus in his lacking the conative capacity to form the belief that p.

First, an indifferent actor's conative incapacity may result from a constitutive lack of sufficient concern for the interests and well-being of others, and at the extreme, a complete lack of such concern.[88] In the latter case, the indifferent actor begins to look either like a psychopath, who cares for himself but has zero concern for others,[89] or like someone suffering from

87. *See, e.g.*, DUFF, *supra* note 55, at 157 (culpable negligence is "essentially a matter . . . of a kind of 'practical indifference.'"); MORAN, *supra* note 83, at 258 ("[T]he indifference account places its focus on the *attitude* displayed by any particular action."); SAMUEL H. 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."); Tadros, *supra* note 59, at 229 (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"); Horder, *supra* note 55, at 501 ("The subjective element in indifference lies . . . in an uncaring attitude towards the victim's relevant protected interests."); Samuel H. Pillsbury, *Crimes of Indifference*, 49 RUTGERS L.J. 105, 151 (1996) ("The key to culpability for failure to perceive is why the person failed to perceive."); Simons, *Culpability and Retributive Theory*, *supra* note 46, at 388 ("Culpable indifference . . . is a desire-state reflecting the actor's grossly insufficient concern for the interests of others."); Simons, *Bad Character*, *supra* note 46, at 264 ("[One] possible culpable indifference standard . . . asks what the actor would have done if he had had a different belief about the relevant risks."); Simons, *Mental States*, *supra* note 57, at 487 ("[R]eckless indifference . . . [means] caring much less about the result than the actor should."); *cf.* Montmarquet, *Culpable Ignorance*, *supra* note 69, at 43 (arguing that an actor is culpable for possessing a belief if and when that belief has been "formed with (or characterized by) an intellectually irresponsible attitude (i.e., an attitude, broadly put, of insufficient regard for truth and evidence)"); George Sher, *Out of Control*, 116 ETHICS 285, 298 (2006) ("[O]ne obvious way to articulate a conception of control that can be instanced even by agents who do not know that they are acting wrongly is to say that someone exercises such control whenever his not believing that he is acting wrongly is explained not by his lack of access to the facts that make his act wrong but rather by some subset of the other beliefs (desires, attitudes, etc.) that make him the person he is.").

88. *See, e.g.*, DUFF, *supra* note 55, at 155 (seeming to equate "feel[ing] indifferent to . . . the safety" of others with "feel[ing] *no* concern for their safety" (emphasis added)); Horder, *supra* note 55, at 503 ("The strongly indifferent person . . . is moved *only* by his own agent-relative values (and prudential reasons relating thereto)." (emphasis added)).

89. *See, e.g.*, RONALD D. MILO, IMMORALITY 61 (1984) ("Perhaps the most distinguishing characteristic of the psychopath is that he is entirely egocentric, without any concern whatsoever for the needs and interests of others."). Writers disagree on whether a psychopath's complete lack of concern for others renders him incapable of knowing or appreciating the difference between right and wrong, or simply incapable of being motivated to act in accordance with his judgments of right and wrong, and on whether the psychopath's condition, however characterized, defeats (in whole or in part) the responsibility he would otherwise bear for his actions. *See, e.g.*, JEFFRIE G. MURPHY, *Moral Death: A Kantian Essay on Psychopathy*, *in* RETRIBUTION, JUSTICE, AND THERAPY 128, 130 (1979) ("[P]sychopaths know, in some sense, what it means to wrong people, to act immorally, [but] this kind of judgement has for them no motivational component at all."); Walter Glannon, *Psychopathy and Responsibility*, 14 J. APPLIED PHIL. 263, 273 (1997) ("[P]sychopaths are deficient in the deep moral knowledge of right and wrong required to be completely responsible for their behaviour, [but] they nonetheless have sufficient moral understanding to be partly responsible for their actions."); Ishtiyaque Haji, *The Emotional Depravity of Psychopaths and Culpability*, 9 LEGAL THEORY 63, 82 (2003) (arguing that the culpability of a psychopath whose "emotional impairment"

profound depression, who has zero concern, not only for others, but for himself as well.[90] Such an actor lacks sufficient desire, or perhaps any desire, to regard others with the concern to which their status as a moral agent entitles them to be treated and thus lacks the conative capacity needed to form the belief that *p*. Moreover, inasmuch as desires cannot directly be willed into existence or willed to be stronger, the indifferent actor so portrayed can do nothing then and there to repair the unfortunate state of his conative constitution.

Second, an indifferent actor's conative incapacity may result from a non-constitutive lack of sufficicnt concern for the interests and well-being of others. Such an actor is not constitutionally indifferent. On the contrary, he ordinarily regards other people as moral agents worthy of respect and is motivated accordingly. The problem is that his own concerns or interests eclipse, outweigh, or trump his concern for others, such that the net effect is to place him in a conative state in which his concern for others is insufficient to support the formation of the belief that *p*.[91] An actor might routinely or habitually find himself in such a state, in which case wc might say that he is an indifferent *person*; or he might only occasionally find himself in such a state, in which case we might say simply that he has an indifferent *attitude* on those occasions.[92] Either way, at the point in time at which the actor should

---

makes an act "difficult to perform," but whose action otherwise satisfies the conditions of responsibility, is not "subvert[ed]" but may be "attenuat[ed]"); Ishtiyaque Haji, *On Psychopaths and Culpability*, 17 LAW & PHIL. 117, 139–40 (1998) ("[I]n a wide range of cases, the problem of attributing responsibility to the psychopath turns essentially on epistemic requirements of responsibility[,] . . . [such that] in a host of primary cases, the psychopath is not to blame for her offenses.").

90. The Latin term for the condition I have in mind here is *accidie*. The term "listlessness" also captures the phenomenon. *See* Alfred R. Mele, *Internalist Moral Cognitivism and Listlessness*, 106 ETHICS 727, 734 (1996) ("In its most severe forms, . . . listlessness [or *accidie*] consists in the total absence of motivation to engage in activities of kinds that formerly were matters of deep personal concern. The phenomenon has been linked . . . to clinical depression."). Christian theologians used the term "sloth" to describe this condition, although that term has come to be associated with laziness. *Compare* 3 ST. THOMAS AQUINAS, SUMMA THEOLOGICA, Pt. II-II, Q. 35, Art. 1, at 1339 (Fathers of the English Dominican Province trans., Christian Classics 1981) (1948) (defining sloth as "an oppressive sorrow, which . . . so weighs upon man's mind, that he wants to do nothing"), *with* WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY, UNABRIDGED 2146 (1986) (listing laziness as a representative synonym).

91. *See, e.g.*, Horder, *supra* note 55, at 502 ("[F]or the weakly indifferent person the existence (and strength) of agent-neutral reasons may be obscured by his or her attachment to the pursuit of agent-relative values in a particular set of circumstances.").

92. *See, e.g.*, *id.* at 505–06 (emphasizing the distinction between "someone [who] is indifferent by *nature* (a claim about that person's character) . . . [and someone whose] particular *action* manifests indifference (a claim about that person's conduct)"); Smith, *supra* note 69, at 555–56 (distinguishing between character traits and "undesirable configuration[s] of desires and aversions that generates the choice to perform that action"); Sverdlik, *supra* note 69, at 144 ("[P]ure negligence often gets explained by the way that various morally significant psychological states affect a person's thought[s]. I am inclined to include in these states both character traits and attitudes.").

have formed the belief that *p*, his insufficient concern for others prevented him from forming it.

So understood, the hypothetical-choice theory maintains that an indifferent actor can fairly be punished, not because he chose to do anything the law prohibits, but simply for being in or possessing a particular conative state, where that state then prevented him from forming the belief that *p*. The objection to any such theory is straightforward. If an indifferent actor could not have formed the belief that *p* because he lacked sufficient concern for others, then however much we might censure him for being so constituted, he has not yet chosen to do anything for which he can fairly be punished. He did of course fail to form the belief that *p*, and as a result, he unwittingly created a lethal risk. Yet he did not choose not to form that belief. On the contrary, he failed to form that belief because he lacked the capacity to form it, and incapacity ordinarily excuses.[93]

Here is the point at which the idea of hypothetical choice figures into the theory. Faced with this objection, that an indifferent actor who fails to form the belief that *p* has not chosen to do anything for which he can fairly be punished, the hypothetical-choice theorist turns, not to prior choice, but to hypothetical or counterfactual choice.[94] Although an indifferent actor has not in fact chosen to impose a lethal risk, he can nonetheless fairly be punished, so the argument goes, if he *would have* chosen to impose such a risk had he been aware of it. Thus if an actor who unwittingly creates a lethal risk *would have* taken that risk had he been aware of it, then he can on this theory fairly be subject to retributive punishment. Indeed, he can fairly be subject to punishment not only for involuntary manslaughter but perhaps for reckless homicide.[95]

---

93. If the hypothetical-choice theorist were at this point to insist that the actor is nonetheless responsible for his ignorance if and because he culpably chose to cause the motivational incapacity giving rise to it, then the hypothetical-choice theory would collapse into the broad prior-choice theory. *See supra* notes 71–82 and accompanying text.

94. Alan Michaels endorses the use of hypothetical choice only when an actor chooses to act despite believing his act would create a less-than-certain risk of causing a statutorily-proscribed harm and the hypothetical inquiry is whether the actor would have chosen to act even if he believed his act would be practically certain to cause the statutorily-proscribed harm. Michaels, *supra* note 48, at 961–62. He refuses to endorse the use of hypothetical choice when an actor does not choose to take such a risk because he is unaware of any such risk and the hypothetical inquiry is whether the actor would have chosen to take the risk if he had been aware of it. *Id.* at 962 n.26. Michaels declines to extend the hypothetical-choice theory in this way in part because he believes that doing so would make the actor's indifference a "mere unconnected thought," as opposed to an "actual culpable mental state linked to an act." *Id.* In my view, this reservation applies with equal force to the circumstances in which Michaels endorses the use of hypothetical choice. A hypothetical choice is not an actual choice no matter how you cut it. According to Michaels, "[i]f the actor is *willing* to endure the result or the circumstance that is a necessary adjunct to his conduct, then he 'accepts' the result or circumstance." *Id.* at 962 (emphasis added). But unless the actor believes the result will occur or that the circumstance exists, he has not yet shown himself "willing to endure" that circumstance or result. He might be willing, but then again, he might not.

95. Horder describes an actor who, had he cared enough, would have realized the risk and not taken it as "weakly indifferent," and an actor who, had he cared enough, would have realized the

This resort to hypothetical choice is ultimately unpersuasive. In some cases we can of course be quite confident about what an actor who was unaware of a risk would have done if he had been aware of it. Perhaps, as has been suggested, the hypothetical-choice theory should be limited to those cases in which the available evidence about what an actor would have done is truly compelling.[96] For example, suppose the actor experiences no genuine regret nor honest remorse when he realizes that his unwitting risk imposition has caused someone's death. Or suppose he actually confesses after the fact that he would have done the same thing even if he had been aware of the risk beforehand. Who would feel sorry punishing such an actor? Aren't we pretty sure that he would have done the same thing even if he had been aware of the risk? Why *not* punish him?

Such cases test our commitment to the principle that punishment should be imposed only for what we choose to do, and not on predictions, however firm, about what we probably would have chosen to do if we possessed beliefs we actually lacked.[97] If we opt to impose punishment in such cases, we should be honest with ourselves about what the actor is being punished for. He is being punished for lacking sufficient concern for others, which lack of concern in turn prevented him from realizing the risk he was creating, and on the basis of which we predict he would have taken the risk even if he had been aware of it. In other words, he is being punished for having desires he should not have had, or for lacking desires he should not have lacked, and as such, he is being punished for who he is, or at least for who he was at the moment, and not for what he has chosen to do.

An actor is responsible for the desires he possesses because his desires, together with his beliefs, constitute who he is. But the question is not merely whether an actor is *responsible* for the desires resulting in his indifference (he is); nor whether he should feel ashamed for being so constituted (he

---

risk but taken it all the same as "strongly indifferent." Horder, *supra* note 55, at 502–03. What Horder calls "weak indifference" and "strong indifference" correspond to what Simons calls "mild indifference" and "modest indifference." Simons, *Bad Character*, *supra* note 46, at 264 n.93. A weakly or mildly indifferent actor would presumably be punished for involuntary manslaughter, whereas a strongly or modestly indifferent actor would presumably be punished for reckless homicide. *See, e.g.*, *id.* at 286 ("Some . . . commentators (including this author) would indeed extend the [cognitive counterfactual] approach to a subcategory of culpably unaware . . . actors: such actors *might be* punished the same as cognitively reckless actors . . . ." (emphasis added)).

96. *See, e.g.*, *id.* at 287 n.130 (identifying six conditions or criteria under which we can have considerable confidence in our predictions). Simons would apparently permit the state to punish an actor who unwittingly creates a lethal risk only where it can be shown that the actor actually believed at time t_1 that he would be creating a lethal risk if he were to φ or fail to φ at time t_2, but no longer possesses that belief, or is no longer aware of that belief, when at time t_2 he actually φs or fails to φ, and where the interval between t_1 and t_2 is "very short." *See id.* Simons describes this set of circumstances as a "deflationary scenario" because the actor possesses the relevant belief at time t_1 but then forgets that belief at time t_2. The actor's mental state therefore "deflates" between time t_1 and t_2. *See id.*

97. *See, e.g.*, *id.* at 280–81 n.121 (noting that the hypothetical-choice theory "is in tension with autonomy values" but maintaining that some compromise of those values is nonetheless worth making).

should); nor even whether he can fairly be reproached for being so constituted (he can be). The question is whether a liberal state can fairly punish him—censure and condemn him through hard treatment[98]—for the desires he possesses if and when those desires happen to render him incapable of forming a belief he should have formed, and in fact, would have formed but for those desires.[99] Although proponents of the hypothetical-choice theory may well disagree, in my view the answer is no. An illiberal state is free so to punish, but a liberal state is not, at least not if it remains true to the commitments defining it.

Can the hypothetical-choice theory explain why one might think retributive punishment is warranted for Sam and Tiffany, but not for the Williamses? Yes, but the explanation it gives is objectionable. It would punish Sam and Tiffany because the desire blinding them to the truth was the desire to climb higher on the social ladder, and it would withhold punishment from the Williamses because the desire blinding them to the truth was their love for their child and the fear of losing him. Sam and Tiffany are therefore punished because the desire blinding them is judged blameworthy, while the desire blinding the Williamses is judged not blameworthy. But again, a liberal state should not punish an actor just because he was unlucky enough to find himself in a situation in which his possession of a blameworthy desire blinded him to a truth he would otherwise have seen.

III. The Culpable Failure of Doxastic Self-Control

If retributive punishment is warranted only if an actor is culpable, if culpability consists in choosing to do wrong, and if no actor guilty of involuntary manslaughter chooses to do wrong, then punishment for involuntary manslaughter is either unwarranted, or the culpability of involuntary manslaughter must consist in something other than the culpability of choice. But if the culpability of an actor who unwittingly creates a lethal risk is not the culpability of choice, then what is it the culpability of?

On the account offered here, the answer is failed self-control.[100] On this account, an actor can fairly be held liable for involuntary manslaughter if and

98. *See* JOEL FEINBERG, *The Expressive Function of Punishment*, *in* DOING AND DESERVING 95, 98 (1970) (arguing that "both the 'hard treatment' aspect of punishment and its reprobative function must be part of the *definition* of legal punishment").

99. *See, e.g.*, Robert Adams, *Involuntary Sins*, 94 PHIL. REV. 1, 21–22 (1985) ("[P]eople may rightly be reproached (and therefore blamed) for involuntary as well as for voluntary moral faults, [but] only voluntary acts and omissions are rightly *punished* by the state."); Angela M. Smith, *Responsibility for Attitudes: Activity and Passivity in Mental Life*, 115 ETHICS 236, 271 (2005) (concluding that we are responsible for our intentional mental states but questioning "whether we are open to the very same kinds of appraisals for [such states] as we are for our voluntary actions").

100. H.L.A. Hart seems to have had a similar idea in mind when he wrote:

> In . . . [some] cases, exemplified in 'provocation' and 'diminished responsibility', if we punish at all we punish *less*, on the footing that, though the accused's capacity for self-control was not absent its exercise was a matter of abnormal difficulty. *He is punished*

when he fails to exercise self-control, and more specifically, if and when he fails to exercise *doxastic* self-control, at least when the actor has the capacity to exercise such self-control and when its exercise is otherwise fair to expect. What follows can be seen as an effort to refine and defend this basic account.

A. The Capacity for Doxastic Self-Control

Hart argued long ago that punishment was warranted only if an actor had the capacity and a fair opportunity to conform his conduct to the requirements of law.[101] If an actor could have done as the law demanded he do *given his capacities*,[102] then the state would be warranted in punishing him for failing to conform to those demands, assuming he had a fair opportunity so to conform. Conversely, punishing him would be unfair if he lacked either the capacity to conform or a fair opportunity to do so, in which case his failure to conform should be excused.[103] If the actor has no such excuse, and if we ask why he failed to exercise his capacity to obey, one immediate answer would be that he chose not to obey. He chose instead to disobey. But one can choose to disobey in two different ways.

First, an actor can choose to disobey not only in the minimal sense that his will executes a desire prompting him to act in a manner inconsistent with the law's demands to refrain from so acting, but in the more robust sense that his will endorses or identifies with that desire. He thereby makes that desire his own. His choice is wholehearted.[104] We can describe such an actor as

---

> *in effect for a failure to exercise [self-]control; and this is also involved when punishment for negligence is morally justifiable.*

HART, *supra* note 68, at 153 (second emphasis added). Michael Moore has endorsed Hart's basic explanation for the culpability of negligence. *See* MOORE, *supra* note 62, at 591 ("Since [the culpability of negligence] is neither the culpability of choice nor of character, it must be a distinct form of culpability, what Hart aptly termed the culpability of unexercised capacity."). In contrast, Sanford Kadish has argued that "our law['s]" refusal to endorse Hart's approach, which Kadish describes as an "excuse of incapacity to know better," is "understandable" in "view of both the practical and moral complexities [such an excuse would] raise[]." Sanford H. Kadish, *Excusing Crime*, 75 CAL. L. REV. 257, 278 (1987).

101. *See* HART, *supra* note 68, at 152, 154–55 (offering various formulations of this general principle); *see also* A.D. Woozley, *Negligence and Ignorance*, 53 PHIL. 293, 301 (1978) ("[A] man would not be found guilty of having acted negligently, if he had no fair opportunity of avoiding it."). Hart's principle has been described as setting forth a "minimal condition of [criminal] liability." R.A. Duff, *Choice, Character, and Criminal Liability*, 12 LAW & PHIL. 345, 347 (1993).

102. HART, *supra* note 68, at 154.

103. This statement reflects what is commonly described as the choice theory of excuse. *See, e.g.*, MOORE, *supra* note 62, at 549–62.

104. *See, e.g.*, MICHAEL E. BRATMAN, *Identification, Decision, and Treating as a Reason*, *in* FACES OF INTENTION 185, 193 (1999) ("[I]dentification is a kind of decision about our desires."); Harry Frankfurt, *Identification and Wholeheartedness*, *in* RESPONSIBILITY, CHARACTER, AND THE EMOTIONS 27, 38 (Ferdinand D. Schoeman ed., 1987) (explaining that an actor acts wholeheartedly on the basis of a desire when that desire is "incorporated into himself by virtue of the fact that he has it *by his own will*").

having defied the law. He has, as Michael Moore has said, "not even tr[ied] to get it right."[105] His culpability is the culpability of defiant choice.

Second, an actor can, as in the case of defiant choice, choose to disobey in the minimal sense that his will executes a desire prompting him to act contrary to the law's demands, but unlike in the case of defiance, the actor's will does not lend its endorsement to that desire or otherwise identify with it. The actor fails to conform his conduct to the law's demands, not because he defies the law, but rather because he fails successfully to resist the desire prompting him to disobey. His will refuses to endorse the desire prompting disobedience, but it nonetheless ends up doing desire's bidding. He fails successfully to exercise self-control. His culpability is thus not the culpability of defiance, but of weakness of will. Defiance and weakness are different forms of the culpability of choice.[106]

The culpability of unwitting risk-creation likewise consists in the failure to exercise self-control. However, failing to exercise self-control can result in an actor choosing to do something the law forbids *or* in his either failing to form a belief the law (implicitly) demands he form or in his failing to become aware of a belief he has already formed. In other words, self-control can be either practical (directed toward action) or doxastic (directed toward belief). Insofar as the culpability of involuntary manslaughter rests on an actor's failure to exercise doxastic self-control, the actor is assumed to have available to him at time t_1 beliefs sufficient to support the formation of the belief that p, as well as the cognitive capacity needed to form that belief. He is also assumed to possess a standing concern for others sufficient to support the formation of the belief that p. The problem is that countervailing desire causes the actor then and there to lack the conative capacity needed to form that belief, or to become aware of it if he has already formed it. Consequently, he fails to form that belief or fails to become aware of his having formed it.

If an actor's conative state at the time he ought to have formed the belief that p renders him unable to form or become aware of that belief, and if ought implies can, then an excuse would normally be in order. Yet inasmuch as the actor also possesses the capacity for doxastic self-control, he has the capacity to alter his conative state. He can resist the desire blocking the belief's formation or his becoming aware of it. If he succeeds he will, without more, either form the belief that p or become aware of it if he has already formed it. In other words, his capacity for doxastic self-control enables him to remove the conative incapacity otherwise preventing the belief from forming or from surfacing into awareness.

An actor's failure to exercise doxastic self-control is not reducible to any prior choice, nor therefore is the theory described here reducible to the

105. MOORE, *supra* note 62, at 590.

106. *See* Stephen P. Garvey, *Passion's Puzzle*, 90 IOWA L. REV. 1677, 1727–29 (2005) (describing the distinction between defiance and weakness in more detail).

prior-choice theory.[107] First, insofar as the prior-choice theory presupposes that the choice the actor makes is a choice to do something (or not do something when the actor is under an obligation to do something), it presupposes that the relevant work of the will is (at least) to execute a desire resulting in an action (or inaction), which action results in turn in the actor's subsequent incapacity to form the belief that p. In contrast, the exercise of doxastic self-control requires the will to perform no such work. Nor does it require the will directly to form the belief that p. Indeed, though the will can accept or reject the beliefs an actor forms,[108] it cannot directly form them. We cannot will to believe.[109] Instead, doxastic self-control requires the will to alter the actor's existing configuration of desires so as to enable the belief that p to form or to enable that belief to reach awareness.[110]

---

107. References to the prior-choice theory in this and the following paragraph are references to the broad prior-choice theory, according to which the actor makes a culpable choice at time t_1 the result of which is an incapacity at time t_2 to form the belief that p.

108. *See, e.g.*, L. JONATHAN COHEN, AN ESSAY ON BELIEF AND ACCEPTANCE 22 (1992) ("Acceptance, in contrast with belief, occurs at will."); Shute, *supra* note 47, at 192 ("Acceptances . . . engage with the will in a different way [than beliefs]. Beliefs are 'passive'. They cannot be acquired directly through an act of will. . . . In contrast, acceptances are 'active': they do respond to the will."). *But cf.* Raimo Tuomela, *Belief Versus Acceptance*, 2 PHIL. EXPLORATIONS 122, 136 (2000) ("[A]cceptance 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.").

109. The claim that we cannot will to believe can be interpreted as a conceptual truth or as a contingent one. For arguments in support of the conceptual-truth interpretation, see BERNARD WILLIAMS, *Deciding to Believe*, *in* PROBLEMS OF THE SELF 136, 148 (1973), in which it is argued that "it is not [merely] a contingent fact that I cannot bring it about, just like that, that I believe something . . . ," and Dion Scott-Kakures, *On Belief and the Captivity of the Will*, 53 PHIL. & PHENOMENOLOGICAL RES. 77, 77 (1993), which defends "a version of the conceptual impossibility claim." For an argument in support of the contingent-truth interpretation, see William P. Alston, *The Deontological Conception of Epistemic Justification*, 2 PHIL. PERSPECTIVES 257, 263 (1988), in which Alston contends that "we are not [in fact] so constituted as to be able to take up propositional attitudes at will."

Other writers argue that we can will to believe. *See, e.g.*, Sharon Ryan, *Doxastic Compatibilism and the Ethics of Belief*, 114 PHIL. STUD. 47, 70 (2003) ("If you have compatibilist intuitions, you should deny [the] premise [that doxastic attitudes are never under our voluntary control]."); Matthias Steup, *Doxastic Voluntarism and Epistemic Deontology*, ACTA ANALYTICA, Summer 2000, at 25, 26 ("If 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."). In any event, accepting the claim that we cannot will to believe does not necessarily force one to reject the view that epistemic justification should be understood as a matter of forming one's beliefs in accordance with one's epistemic duties. *See, e.g.*, Robert Audi, *Doxastic Voluntarism and the Ethics of Belief*, *in* KNOWLEDGE, TRUTH, AND DUTY 93 (Matthias Steup ed., 2001) (arguing that the lack of voluntary control over belief-formation "do[es] not prevent our sustaining a deontic version of an ethics of belief"); Richard Feldman, *Voluntary Belief and Epistemic Evaluation*, *in* KNOWLEDGE, TRUTH, AND DUTY, *supra*, at 77, 90 ("[D]eontological judgments about belief . . . do not imply that belief is voluntary."). *But see* Alston, *supra*, at 294 ("[W]e are ill advised to think of epistemic justifieation in terms of freedom from blame for believing.").

110. *Cf.* Jeanette Kennett & Michael Smith, *Synchronic Self-Control Is Always Non-Actional*, 57 ANALYSIS 123, 128 (1997) (claiming that "all exercises of synchronic self-control are non-actional" but not "that all exercises of non-actional self-control are synchronic").

Second, the prior-choice theory presupposes that the actor's choice *causes* his incapacity and thus his ignorance. The choice is prior to the ignorance and the cause of it. The prior-choice theory is therefore a diachronic theory. Indeed, one of the prior choices an actor might make is the choice not to exercise diachronic self-control. Imagine, for example, that Ulysses had chosen not to tie himself to the mast, thereby leaving himself unable later to resist the Sirens' song. In contrast, the exercise of doxastic self-control is synchronic, not diachronic. The failure to exercise such self-control does not cause the actor's ignorance. On the contrary, because the actor already has internal access to all the evidence he needs to form the conscious belief that *p*, the successful exercise of doxastic self-control causes the actor's ignorance to lift, while the failure to exercise such self-control allows it to endure. Far from being the cause of his ignorance, the actor's will is the potential solution to it.

B. A Fair Opportunity to Exercise Doxastic Self-Control

Before an actor can fairly be punished for failing to exercise doxastic self-control, he must not only have the capacity for such self-control, he must also have a fair opportunity to exercise that capacity. At some point the law's demand to exercise doxastic self-control will become an unfair demand, and when that point is reached, an actor's failure to exercise such self-control should be excused.

Reasonable minds can disagree on the circumstances under which the law can and cannot fairly demand the exercise of doxastic self-control.[111] Having said that, I would submit that the law can fairly demand the exercise of doxastic self-control if and when 1) the actor either suspects that *p* but does not yet believe that *p* (a case of nonwillful ignorance) or believes that *p* but is unaware of that belief (a case of self-deception); and 2) the desire preventing the formation of the belief that *p*, or preventing the actor from becoming aware of that belief, is a desire the influence of which the actor should have controlled. The first condition is cognitive (related to belief); the second, conative (related to desire). If both of these conditions obtain, then the actor's failure to exercise doxastic self-control is culpable, and he can fairly be subject to retributive punishment. If either fails to obtain, then the actor's failure to exercise such self-control is nonculpable, and he should be excused.[112]

---

111. *See, e.g.*, Jeremy Horder, *Criminal Culpability: The Possibility of a General Theory*, 12 LAW & PHIL. 193, 202–03 (1993) ("[The] key distinguishing mark [of what Horder calls the capacity theory of criminal culpability], the insistence on the fairness of opportunities to avoid wrongdoing (at least where morally stigmatic crimes are in issue) is ambiguous just where it matters, namely over the question of what is 'fair' or 'unfair' about an opportunity to avoid wrongdoing.").

112. Insofar as the beliefs available to an actor at time t_1 support the formation of the further belief that the actor's conduct is creating a risk of causing death, the actor's failure to form that belief is never in the law's eyes permissible or justifiable.

1. The Cognitive Condition.—According to one account, an actor who fails to exercise doxastic self-control can fairly be punished if, all else being equal, his available background beliefs are sufficient to support the formation of the conscious belief that p.[113] But the fact that an actor's background beliefs are sufficient to support the formation of the conscious belief that p, and indeed, that the actor would have formed that belief but for the existence of some countervailing desire preventing its formation, is not yet enough to make it fair to punish him for failing to control that desire. For although his background beliefs should perhaps alert him to the need to exercise the self-control it would take to form the conscious belief that p, they do not without more in fact alert him to that need. Something more than sufficient background beliefs is required.

According to another account, an actor who fails to exercise doxastic self-control so as to permit the formation of the belief that p can fairly be punished if, all else being equal, the actor believes that he is imposing a risk of causing some harm short of death.[114] He might realize his conduct is

---

113. Oliver Wendell Holmes appears sympathetic to this position when he writes in *The Common Law*:

> It is enough [for criminal liability] that . . . circumstances [making the conduct dangerous] were actually known as would have led a man of common understanding to infer from them the rest of . . . the present state of things. For instance, if a workman on a house-top at mid-day knows that the space below him is a street in a great city, he knows facts from which a man of common understanding would infer that there were people passing below. He is therefore bound to draw that inference . . . . If then, he throws down a heavy beam into the street, he does an act which a person of ordinary prudence would foresee is likely to cause death, or grievous bodily harm, and he is dealt with as if he foresaw it, whether he does so in fact or not.

OLIVER WENDELL HOLMES, JR., THE COMMON LAW 46–47 (Little, Brown & Co. reprint) (Mark D. Howe ed., Harvard Univ. Press 1963) (1881); *see also* George P. Fletcher, *The Theory of Criminal Negligence: A Comparative Analysis*, 119 U. PA. L. REV. 401, 426 (1971) ("[I]nadvertence to a risk as well as choosing to take a particular risk might warrant the just censure of others. When the circumstances give the actor reason to think that his conduct risks harm to another, his failure to apprise himself of the risks latent in his conduct is culpable."). Steven Sverdlik likewise appears sympathetic when he writes:

> It might be objected . . . that when an act of pure negligence occurs [i.e., an act of negligence not attributable to any prior culpable choice], the agent had no *reason* to reflect morally, or therefore, to act any differently from the way she did. This, in turn, would make the purely negligent agent seem inappropriate as a target of blame. This is surely mistaken. If our principles or norms give us reason to do things, once we see that they apply, they surely must give us a reason to find out if they in fact apply. A morally conscientious agent who is committed to conforming her behavior to moral norms, in other words, has a reason to find out if she is likely to violate one of them.

Sverdlik, *supra* note 69, at 142. In other words, according to Sverdlik, when the background beliefs available to an actor at time t_1 support the formation of the belief that p, such that the actor need only "reflect morally" on those background beliefs in order to form that belief, an actor who fails so to reflect is culpable, because a "morally conscientious" agent always and everywhere has reason so to reflect.

114. Victor Tadros might be interpreted as being at least sympathetic to this position when he writes:

> [An actor's] failure to investigate the risks where it is appropriate to do so shows only that he has an awareness that his activity is risky to a degree and in a way that *ought to*

dangerous but not realize just how dangerous. He might see a risk of some harm but no risk of death. On this account, his awareness of the lesser risk makes it fair to demand the exercise of doxastic self-control. An actor who believes that he is imposing a nonlethal risk and who takes that risk without good reason could of course fairly be punished for choosing to take that risk. Nonetheless, though his belief that he is imposing such a risk should perhaps alert him to the need to exercise the self-control it would take to form the conscious belief that *p*, it does not without more in fact alert him to that need. Again, something more is needed.

When would it be fair to punish an actor for failing to exercise doxastic self-control? In my view it would, all else being equal, be fair when the actor possesses one of two cognitive states. The first can be described as a state of nonwillful ignorance in which the actor suspects, but does not yet believe, that *p*.[115] The second can be described as a state of self-deception in which the actor unconsciously believes that *p* but sincerely avows ignorance of *p*.[116] An actor who suspects that *p*, or who unconsciously believes that *p*, not only should be aware of the need to exercise doxastic self-control. He is in fact aware of that need. Such awareness is internal to each of these cognitive states.[117]

When an actor's background beliefs are sufficient to support the formation of the belief that *p*, but the actor nonetheless only suspects that *p* because desire keeps him from forming the belief that *p*, the actor's suspicion will nag at his consciousness until an exercise of self-control allows his suspicion to blossom into belief or until desire prevails and dispels the suspicion altogether. Similarly, if the actor actually forms the belief that *p*, but desire simultaneously pushes that belief into the unconscious, that unconscious belief will nag at the actor's consciousness until an exercise of self-control releases it into consciousness or until desire represses it beyond the reach of (ordinary) self-control altogether.

a. Nonwillful Ignorance.—Because nonwillful ignorance is related to the better-known doctrine of willful ignorance (also known as willful blindness), one way to explain the former is to begin with an explanation of

---

lead him to investigate. He need not be aware that there is a risk that the *actus reus* of an offence would come about, let alone that it be the *actus reus* of the offence with which he is charged.

TADROS, *supra* note 55, at 257 (initial emphasis added).

115. *See infra* notes 118–30 and accompanying text (discussing nonwillful ignorance).

116. *See infra* notes 131–43 and accompanying text (discussing self-deception).

117. It follows from this analysis that an actor who was so self-absorbed or indifferent to the well being of others that he never even formed the suspicion that *p* or the unconscious belief that *p* should not be liable for involuntary manslaughter. If this result is perverse, it seems to me that such perversity is the price one pays for a criminal law that refuses to punish people for being unlucky enough to find themselves in a situation in which their self-absorption or indifference prevents them from forming beliefs they would have formed had they not been so self-absorbed or indifferent.

the latter.[118] Willful ignorance is a familiar though controversial doctrine.[119] In its most general formulation the doctrine provides that when an offense requires knowledge with respect to the existence of an attendant circumstance (call this circumstance p^*)—for example, that the suitcase he is carrying contains a controlled substance—but the actor lacks such knowledge, the actor will nonetheless be treated under certain circumstances as if he had such knowledge, even though he did not.

Scholars have offered at least three accounts of willful ignorance. These accounts differ because each begins with a different analysis of the concept of knowledge. The first account begins with the standard philosophical analysis of knowledge, according to which knowledge is "justified true belief."[120] On this account, the knowing actor and the willfully ignorant actor both believe that p^* (and it is true that p^*), but whereas the knowing actor's belief is justified based on the evidence internally available to him, the willfully ignorant actor's belief is not.[121] Nonetheless, the willfully ignorant actor could easily have gathered the additional evidence needed to transform his unjustified belief into a justified belief (and thus into knowledge) had he chosen to do so. Insofar as he chose not to do so because he wanted for no good reason not to know that p^*, the law treats him as if he did know.

---

118. The wording of willful blindness instructions in the federal courts varies among the circuits. United States v. Alston-Graves, 435 F.3d 331, 338 (D.C. Cir. 2006). For one influential case discussing willful ignorance, see United States v. Jewell, 532 F.2d 697, 704 (9th Cir. 1976) (en banc). The Model Penal Code's definition of willful ignorance is contained in § 2.02(7), which provides that "[w]hen knowledge of the existence of a particular fact is an element of an offense, such knowledge is established if a person is aware of a high probability of its existence, unless he actually believes that it does not exist." MODEL PENAL CODE § 2.02(7) (1962). For discussions of the doctrine in English law, see ASHWORTH, *supra* note 38, § 5.3(e), at 191–93, and SIMESTER & SULLIVAN, *supra* note 83, § 5.4(i), at 147–49.

119. Academic commentary on the willful ignorance doctrine focuses on three related questions. First, what does it mean to say that an actor is willfully ignorant? Second, assuming an actor who is willfully ignorant with respect to p^* lacks knowledge with respect to p^*, is the culpability of an actor who is willfully ignorant with respect to p^* nonetheless morally equivalent to that of an actor who knows that p^*? Third, assuming the culpability of an actor who is willfully ignorant with respect to p^* is morally equivalent to that of an actor who knows that p^*, does it nonetheless violate the principle of legality for courts to permit liability to be imposed on a willfully ignorant actor when the offense charged requires knowledge and no statutory provision exists defining willful ignorance as "knowledge"? My focus here is on the first question.

120. *See, e.g.*, RICHARD FELDMAN, EPISTEMOLOGY 15 (2003) ("Philosophers often say that what is needed for knowledge, in addition to true belief, is *justification* for the belief."). This definition ignores Gettier problems. *See* Edmund L. Gettier, *Is Justified True Belief Knowledge?*, 23 ANALYSIS 121 (1963). Gettier problems involve cases in which an actor's belief is both justified and true but nonetheless does not constitute knowledge. *See id.* at 121 (arguing that the claim that "justified true belief" constitutes a "*sufficient* condition for the truth of the proposition that S knows that P" is false).

121. *See* Douglas N. Husak & Craig A. Callender, *Wilful Ignorance, Knowledge, and the "Equal Culpability" Thesis: A Study of the Deeper Significance of the Principle of Legality*, 1994 WIS. L. REV. 29, 51 ("[M]any (but not all) wilfully ignorant defendants do not possess knowledge of the incriminating proposition p . . . [because such] defendants . . . lack[] sufficient [internal] justification to know p.").

The second account begins with one reading of the Model Penal Code's § 2.02(2) definition of knowledge.[122] On this account, an actor "knows" that p^* if he is aware that it is practically certain that p^*, whereas he is reckless with respect to p^* if he is aware, not that it is practically certain that p^*, but only that a substantial risk exists that p^*.[123] If an actor who is aware that p^* necessarily believes that p^*, a knowing actor is one who believes that the probability that p^* is practically certain ($p^*_{practically\ certain}$), whereas a reckless actor is one who believes that the probability that p^* is substantial but not practically certain ($p^*_{substantial}$). The difference between a knowing actor and a reckless actor thus turns on the different content of their respective beliefs. An otherwise reckless actor who could have easily gathered the additional information needed to transform his belief that $p^*_{substantial}$ into the belief that

122. The Model Penal Code's definition of willful ignorance, contained in § 2.02(7), suffers from a number of problems. *See, e.g., id.* at 37–39 (describing a "number of grounds" on which § 2.02(7) is "defective"); Michaels, *supra* note 48, at 983 (stating that § 2.02(7)'s "high probability approach to deciding whieh cases short of actual awareness should be treated as equivalent to knowledge has been soundly criticized"); Simons, *Mental States*, *supra* note 57, at 501 n.128 (claiming that § 2.02(7) is "an awkward statement of the wilful blindness doctrine"). For example, the text of § 2.02(7) presupposes that an actor can at the same time believe that the probability that p exists is high while at the same time believing that p does not exist. But the belief that the probability that p exists is high precludes the belief that p does not exist, or at least precludes holding that belief rationally. *See, e.g.*, MICHAEL S. MOORE, LAW AND PSYCHIATRY 86 (1984) (suggesting that such an actor is irrational); Larry Alexander, *Lesser Evils: A Closer Look at the Paradigmatic Justification*, 24 LAW & PHIL. 611, 626 n.34 (2005) ("If someone believes it is more likely than not the case that F exists, then he cannot believe that F does not exist."); Jonathan L. Marcus, Note, *Model Penal Code Section 2.02(7) and Willful Blindness*, 102 YALE L.J. 2231, 2255 (1993) ("[I]t is difficult to imagine how one can simultaneously be aware of a high probability that a fact exists yet believe that it does not exist."). The discussion in the text therefore ignores § 2.02(7).

123. *See, e.g.*, Robin Charlow, *Wilful Ignorance and Criminal Culpability*, 70 TEXAS L. REV. 1351, 1382 (1992) ("[I]n order to 'know' one must be aware of the certainty or near certainty of a fact, and in order to be 'reckless' one must be aware of, at most, the substantial probability of a fact . . . ."); Simons, *Mental States*, *supra* note 57, at 482, 500 (stating that when "recklessness . . . refer[s] to . . . a state of belief" it means a belief that one is creating a substantial risk of causing a proscribed result or a belief that "there is a substantial risk" that an attendant circumstance exists). Larry Alexander has recently argued that the idea of a "reckless *belief*" is "incoherent," at least if one relies on the MPC's § 2.02 definition of recklessness. Alexander, *supra* note 122, at 624. According to Alexander, because an actor is reckless under the Code only if he is "conscious of an unjustified risk[,] [i]n the case of beliefs, presumably, a reckless belief would be one that the defendant both holds and is at the same consciously aware is (too) likely to be untrue." *Id.* at 624–25. Thus, according to Alexander, an actor who recklessly believes that p^* is one who believes that p^* while at the same time believing he should not believe that p^* (because he believes that p^* is likely to be untrue). If so, then the description Alexander gives of an actor who recklessly believes that p^* is that of an actor who *akratically* believes that p^*, i.e., the actor believes that p^* while at the same time believing that he should not (on epistemic grounds alone) believe that p^*. Whether or not such a state is conceptually incoherent, as Alexander suggests, or simply irrational, is a matter on which one can find disagreement. *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 irrational, but incoherent."), *and* David Owens, *Epistemic Akrasia*, 85 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."), *and* Alfred R. Mele, *Incontinent Believing*, 36 PHIL. Q. 212, 217 (1986) (arguing that "full-blown incontinent believing" is possible).

$p^*_{practically\ certain}$ (and thus into knowledge) is willfully ignorant if he chose not to do so because he wanted for no good reason not to know. Under these circumstances the law treats him as if he did know.

The third account begins with a different reading of the Model Penal Code's § 2.02(2) definition of knowledge. On this account, an actor "knows" that p^* if he is aware that p^*. If an actor who is aware that p^* necessarily believes that p^*, an actor who knows that p^* is simply one who believes that p^*. In contrast, an actor who is reckless with respect to p^* is one who *suspects* that p^* but does not yet believe that p^*. The difference between the knowing actor and the reckless actor thus turns on the different cognitive attitude each possesses toward p^*, not on the content of p^*. An otherwise reckless actor who could easily have gathered the additional information needed to transform his suspicion into belief (and thus into knowledge) is willfully ignorant if he chose not to do so because he wanted for no good reason not to know.[124] Under these circumstances the law treats him as if he did know.

Despite their differences, these accounts share a common structure. The willfully ignorant actor lacks knowledge, however that concept is analyzed, but he could have readily *done something* to gather the additional evidence needed to transform his lack of knowledge into knowledge. The actor realizes he could have readily gathered this additional evidence, but chooses for no good reason not to do so. The doctrine of willful ignorance therefore treats a non-knowing actor as if he were a knowing actor when he chooses for no good reason willfully to defy an implicit obligation to act: to gather additional and readily available evidence.[125] The doctrine assumes that the nonknowing epistemic state (however characterized) in which the actor finds himself is that which fairly triggers his duty to act. It also assumes that had the actor discharged his duty to gather additional evidence, his lack of

124. *See* Husak & Callender, *supra* note 121, at 39–40 (arguing that an actor is willfully ignorant with respect to p^* if he suspects that p^*, his suspicion is warranted, he fails to gather additional evidence when it would be easy to do so, and he fails to gather additional evidence because he wants not to believe that p^*).

125. Alan Michaels criticizes this general account of the willful ignorance doctrine, which he calls the "purposeful avoidance" approach, on the ground that it suffers from both overinclusiveness and underinclusiveness. In other words, it characterizes some actors as willfully ignorant when they should not be so characterized (and is thus overinclusive), and it fails to characterize other actors as willfully ignorant when they should be so characterized (and is thus underinclusive). *See* Michaels, *supra* note 48, at 986–87. The account against which Michaels judges the purposeful-avoidance account to be over and underinclusive is what he calls the "acceptance" approach, according to which an actor who acts when he is "merely aware that the circumstance *might* exist," *id.* at 977 (emphasis added), is treated as if he were aware that it *does* exist, *see id.* at 963, if he would still have acted as he did if he had in fact been aware of its existence. *See id.* at 961; *see also* David Luban, *Contrived Ignorance*, 87 GEO. L.J. 957, 974–75 (1999) (proposing a similar counterfactual approach). This approach is objectionable because it punishes an actor not for what he did, but for what he might have done if his cognitive state had been something other than what it actually was. All in all, however, the purposeful-avoidance account of the willful ignorance doctrine seems to me to be the most satisfactory account so far proposed.

knowledge would be transformed into knowledge without any additional action on his part.

Nonwillful ignorance can best be understood by contrasting it to the third account of willful ignorance, according to which the willfully ignorant actor suspects that p^* but does not yet believe that p^*. The willfully ignorant actor does not believe that p^* because he does not have internally available to him evidence sufficient to support the formation of that belief, and chooses for no good reason not to gather the necessary additional evidence, even though he could easily have done so. In contrast, the nonwillfully ignorant actor already has internally available to him all the evidence needed to form the belief that p^*. He fails to believe that p^*, not because he fails to gather more information, but simply because desire prevents him from forming that belief. His conative state permits him to form the suspicion that p^*, but it prevents him from forming the belief that p^*.

Desire can trigger a number of mechanisms that cause an actor to fail to form a belief even though the evidence available to him supports its formation. For example, desire might cause him to misinterpret the available evidence and so see the world, one might say, through rose-colored glasses.[126] It might cause his mind to focus on or attend to favorable evidence and to turn away from unfavorable evidence.[127] It might cause some of the available evidence to be more salient or vivid to him. It might cause him to clutter his mind with other thoughts, thereby crowding out the unwanted belief.[128] Or it might cause him to deploy exclusionary categories, stigmatizing the offending belief as crazy, absurd, ludicrous, and so forth, thereby keeping it from fully forming.[129] Each of these mechanisms operates behind the actor's back, so to speak, without the actor ever being aware of what is going on. Consequently, whereas the willfully ignorant actor makes a conscious choice not to know, because he makes a conscious choice not to gather additional evidence, the nonwillfully ignorant actor makes no such choice. His ignorance is rooted in the covert operations of desire, not in the conscious operation of the will.

A willfully ignorant actor can transform his suspicion into belief only through action. He must gather additional information, confirming his suspicion that p^*, and thereby transforming that suspicion into the belief that

126. *See, e.g.*, Alfred R. Mele, *Real Self-Deception*, 20 BEHAV. & BRAIN SCI. 91, 94 (1997) (discussing negative and positive misinterpretation); *see also* Kent Bach, *An Analysis of Self-Deception*, 41 PHIL. & PHENOMENOLOGICAL RES. 351, 358–60 (1981) (discussing "rationalization").

127. *See, e.g.*, Bach, *supra* note 126, at 360–61 (discussing "evasion"); Mele, *supra* note 126, at 94 (discussing "selective focusing/attending").

128. *See, e.g.*, Bach, *supra* note 126, at 361–62 (discussing "jamming").

129. *See, e.g.*, Kent Bach, *Thinking and Believing in Self-Deception*, 20 BEHAV. & BRAIN SCI. 105, 105 (1997) [hereinafter Bach, *Thinking and Believing*] (discussing "exclusionary categories"); *see also* Kent Bach, *Emotional Disorder and Attention*, *in* PHILOSOPHICAL PSYCHOPATHOLOGY 51, 61–65 (George Graham & G. Lynn Stephens eds., 1994) [hereinafter Bach, *Emotional Disorder and Attention*] (same).

*p**. In contrast, a nonwillfully ignorant actor can transform his suspicion into belief only through doxastic self-control. Rather than gather additional evidence, he need only alter his conative constitution, changing its extant balance of desires. He must "turn off" the desire triggering the mechanisms preventing him from acquiring the belief that *p*. If he succeeds, his suspicion will immediately ripen into belief.

The important point is that the willfully ignorant and the nonwillfully ignorant actor both harbor a suspicion with respect to the relevant fact. This suspicion gives the willfully ignorant actor an internal reason to gather the additional evidence needed to support the formation of the belief that the fact exists, and it gives the nonwillfully ignorant actor an internal reason to exercise the doxastic self-control needed to allow that belief to form. Accordingly, each actor has a fair opportunity, all else being equal, to acquire the belief the law expects him to acquire. The failure to acquire that belief is thus, all else being equal, a culpable failure.[130]

b. Self-Deception.—Being in a state of self-deception also gives an actor an internal reason to exercise doxastic self-control. While the correct analysis of self-deception has been and remains the subject of ongoing dispute,[131] three accounts of what it means to say a person is self-deceived have tended to dominate the discussion. The first two accounts are for different reasons unsatisfactory. The third provides the needed analysis.

The first account analogizes self-deception, in which one and the same person is both deceived and deceiver, to the more familiar case of other-deception, in which one person is the deceived and another is the deceiver.[132]

---

130. As I have construed it here, the willful blindness doctrine provides that if an actor is willfully ignorant as to the existence of an attendant circumstance, the law will treat him as if he did know of its existence, though one can question whether the culpability of a willfully ignorant actor is the same as, or less than, that of a knowing actor. Similarly, an actor who is nonwillfully ignorant as to the existence of a risk might, all else being equal, be fairly treated as if he were aware of that risk, i.e., as if he were reckless, though again, one can question whether the culpability of a nonwillfully ignorant actor is the same as, or less than, that of a reckless actor. It would in any case be more precise to say that the *nature* of their culpability is different: the reckless actor's culpability consists in choosing to impose an unjustified risk, whereas the culpability of the nonwillfully ignorant actor consists in failing to exercise doxastic self-control when the law fairly expects such control.

131. The philosophical literature on self-deception is surprisingly large. For helpful books or collections dedicated to the topic, see ANNETTE BARNES, SEEING THROUGH SELF-DECEPTION (1997); HERBERT FINGARETTE, SELF-DECEPTION (Univ. of Cal. Press 2000) (1969); M.R. HAIGHT, A STUDY OF SELF-DECEPTION (1980); MIKE W. MARTIN, SELF-DECEPTION AND MORALITY (1986); ALFRED R. MELE, SELF-DECEPTION UNMASKED (2001); PERSPECTIVES ON SELF-DECEPTION (Brian P. McLaughlin & Amélie Oksenberg Rorty eds., 1988); SELF-DECEPTION AND PARADOXES OF RATIONALITY (Jean-Pierre Dupuy ed., 1998); SELF-DECEPTION AND SELF-UNDERSTANDING (Mike W. Martin ed., 1985). The works of Alfred Mele and Robert Audi have been most influential on my own thinking.

132. *See, e.g.*, DONALD DAVIDSON, *Deception and Division*, *in* PROBLEMS OF RATIONALITY 199, 207 (2004) ("[S]elf-deception is like lying; there is intentional behaviour which aims to produce a belief the agent does not, when he institutes the behaviour, share."); David Pears, *The*

In cases of other-deception, one actor, knowing the truth that p^* and acting with the intent to deceive the other actor, causes the other actor falsely to believe that not-p^*. Analogizing self-deception to other-deception therefore means imagining that an actor, knowing the truth that p^*, and acting with the intent to deceive himself, causes himself to believe not-p^* at the same time that he believes that p^*. But how, at least in the ordinary case,[133] can an actor get himself to believe a proposition he knows is false? Moreover, how can anyone rationally believe both p^* and not-p^* at the same time? The first account of self-deception thus leads to paradox and contradiction.[134]

The second account tries to explain self-deception without giving rise to these problems. According to this account, an actor is self-deceived if, because he wants to believe that p^*, he forms the belief that p^* when the available evidence provides greater support for the belief that not-p^*.[135] An

---

Goals and Strategies of Self-Deception, *in* THE MULTIPLE SELF 59, 60 (Jon Elster ed., 1986) ("[T]he most extreme form of [irrational belief-formation] is to believe both *p* and *not-p*, which is what is required by the full surface connotation of the word 'self-deception.'"); Ralph Demos, *Lying to Oneself*, 57 J. PHIL. 588, 588 (1960) ("[S]elf-deception entails that B believes both *p* and not-*p* at the same time.").

According to Dion Scott-Kakures, the "internal irrationality" commonly associated with the vernacular conception of self-deception is based, not on the idea that the self-deceived actor simultaneously holds contradictory beliefs, but rather on the idea that he simultaneously believes that p^* (because he wants to believe that p^*) and that he should not (presumably on epistemic grounds alone) believe that p^*. *See* Dion Scott-Kakures, *Self-Deception and Internal Irrationality*, 56 PHIL. & PHENOMENOLOGICAL RES. 31, 44 (1996) ("[T]he internal irrationality of self-deception would appear to be a matter of an agent's coming to believe that not-p (and somehow getting himself to believe it) while believing that he ought not to believe it."). Other writers would probably characterize such an actor as epistemically akratic or incontinent, not self-deceived. *See, e.g.*, ALFRED R. MELE, IRRATIONALITY: AN ESSAY ON AKRASIA, SELF-DECEPTION, AND SELF-CONTROL 112–13 (1987) (setting forth an analysis of "strict incontinent belief").

133. One can of course construct cases in which an actor sets out with the intent to get himself to believe a proposition he knows to be false, and indeed, manages to induce the false belief. But these cases do not represent "garden variety" cases of self-deception. *See* MELE, *supra* note 131, at 16–17 ("Intentionally deceiving oneself is unproblematically possible[, but these] . . . cases are remote from garden-variety self-deception."). Cases of this sort typically depict an actor who, realizing at time t_1 that p^* is false, not only sets out at time t_1 to cause himself to believe that p^* at time t_2. They also depict the actor as setting out to cause himself to forget at time t_2 that he had intentionally induced the belief that p^*, or as one who was aware at time t_1 that he would be likely to forget at time t_2 that he had intentionally induced that belief. *See id.* at 16–17 (describing a "forgetful prankster" who intentionally writes false diary entries in order to deceive himself in the future).

134. In an effort to dispel the paradox, proponents of this account typically postulate that the mind is divided or partitioned in some way such that the desired belief resides in one part of the mind and the undesired belief resides in another part. *See, e.g.*, DAVIDSON, *supra* note 132, at 211 ("[P]eople can and do sometimes keep closely related but opposed beliefs apart. To this extent we must accept the idea that there can be boundaries between parts of the mind . . . ."); Pears, *supra* note 132, at 70 ("[T]he line dividing [the deceiving subject] from [the object of deception] . . . is functional insulation."). Davidson emphasizes that in speaking of the "mind as being *partitioned*, [he] mean[t] no more than that a metaphorical wall separated the beliefs which, allowed into consciousness together, would destroy at least one." Donald Davidson, *Who Is Fooled?*, *in* SELF-DECEPTION AND PARADOXES OF RATIONALITY, *supra* note 131, at 8.

135. This account of self-deception is prominently associated with the work of Alfred Mele. For the most complete defense of this view, see MELE, *supra* note 131. Shorter versions of Mele's

actor is thus self-deceived when, though the evidence available to him does not support the belief that p^*, the actor nonetheless believes that p^* because he wants to believe that p^*. One problem with this account is that it fails to explain the psychic tension an actor experiences when self-deceived. It depicts the self-deceived actor as blissfully ignorant of the truth, with nothing about his cognitive state giving him any reason to question or doubt his belief that p^*. But if some such psychic tension or anxiety is indeed part of the pre-analytic phenomenon of self-deception,[136] then any account leaving it unexplained is, all else being equal, an inadequate account. Indeed, one might say that such an account offers an analysis of self-caused deception[137] or perhaps wishful thinking,[138] but not yet an analysis of self-deception.[139]

---

account are also available. *See generally* Alfred R. Mele, *Emotion and Desire in Self-Deception*, *in* PHILOSOPHY AND THE EMOTIONS 163 (Anthony Hatzimoysis ed., 2003); Mele, *supra* note 123. For what seem to be materially similar accounts, see BARNES, *supra* note 131, at 117, which sets forth necessary and sufficient conditions for "self-deceiving oneself," and Ariela Lazar, *Deceiving Oneself or Self-Deceived? On the Formation of Beliefs "Under the Influence,"* 108 MIND 265, 265 (1999), which argues that "self-deceptive beliefs are direct expressions of the subject's wishes, fears and hopes."

136. *See, e.g.*, ROBERT AUDI, *Self-Deception, Rationalization and the Ethics of Belief*, *in* MORAL KNOWLEDGE AND ETHICAL CHARACTER 131, 152 (1997) ("[Self-deception] tends to be an unstable condition and to exist only so long as there is a balance between the pressure of the evidence . . . and the strength of the defenses maintaining the veil that camouflages that knowledge from consciousness."); Mike W. Martin, *Self-Deceiving Intentions*, 20 BEHAV. & BRAIN SCI. 122, 123 (1997) ("[A]lthough self-deception does not involve fully conscious contradictory beliefs, typically it does involve a cognitive conflict."); Scott-Kakures, *supra* note 132, at 49 ("[T]hough intuitions differ, . . . epistemic tension, instability and fragility seems to me quite fundamental to self-deception."). Mele maintains that any such tension should not be included among the necessary conditions for being self-deceived. MELE, *supra* note 131, at 52–53.

137. *See* Robert Audi, *Self-Deception vs. Self-Caused Deception: A Comment on Professor Mele*, 20 BEHAV. & BRAIN SCI. 104, 104 (1997) (arguing that self-caused deception is different from self-deception).

138. *See* Scott-Kakures, *supra* note 132, at 36–37 (arguing that Mele's purported analysis of self-deception is in fact an analysis of wishful thinking).

139. How one distinguishes wishful thinking (or believing) from self-deception depends on one's analysis of self-deception. Some who adopt the first account described in the text, according to which self-deception consists in simultaneously believing p^* and not-p^*, when the available evidence supports the belief that p^*, maintain that self-deception involves an actor believing that p^*, *either* because he wants to believe that p^*, *or* because (oddly enough) he does not want to believe that p^*, whereas wishful thinking is limited to the former case. Wishful thinking is therefore conceived of as a subset of self-deception. *See, e.g.*, DAVIDSON, *supra* note 132, at 206 (drawing a distinction between wishful thinking, in which the belief is always welcome or positive, and self-deception, in which the belief "may be painful"). *But see* MELE, *supra* note 131, at 4 (describing the former case as a case of "straight" self-deception, and the latter as a case of "twisted" self-deception).

Some who adopt the second account of self-deception described in the text, according to which self-deception consists in believing that p^* because one wants to believe that p^*, when the available evidence supports the belief that not-p^*, maintain that wishful thinking consists in believing that p^* because one wants to believe that p^*, when the available evidence supports neither the belief that p^* nor the belief that not-p^*. *See, e.g.*, MELE, *supra* note 131, at 74 ("[W]ishful thinkers may encounter weaker counterevidence than self-deceivers." (citation omitted)); Bela Szabados, *Wishful Thinking and Self-Deception*, 33 ANALYSIS 201, 205 (1973) (claiming that the self-deceiver "has good grounds" for the belief he does not want to believe, whereas the "wishful thinker does not have such grounds").

The third account answers this objection. On this account, an actor is self-deceived if the evidence available to him not only supports the formation of the belief that p^*, but he in fact forms that belief. The actor is self-deceived because, though he believes that p^*, he holds that belief unconsciously,[140] and he holds that belief unconsciously because he does not want to be aware or conscious of it. The desire not to believe that p^* triggers various mechanisms that prevent the belief from rising to the surface.[141] Accordingly, we sometimes say that the self-deceived actor knows the truth "in his heart."[142] Nonetheless, the self-deceived actor would, if asked, sincerely avow that he possesses no belief that p^*, or perhaps even that he believes that not-p^*. Because he unconsciously believes that p^* while at the same time consciously avowing no belief that p^*, or a belief that not-p^*, the self-deceived actor, unlike the actor in a state of self-caused deception, *does* experience the psychic tension commonly associated with self-deception.[143]

Finally, some who adopt the third account of self-deception described in the text, according to which self-deception consists in simultaneously avowing that not-p^* and unconsciously believing that p^*, when the available evidence supports the belief that p^*, maintain that wishful thinking consists in believing that p^* when the available evidence supports the belief that not-p^* (which is what proponents of the second account would describe as self-deception). *See, e.g.*, COHEN, *supra* note 108, at 144 ("[I]n ordinary wishful thinking we not only . . . desire that p and, as a result, come to think that p: we are all the more able . . . to come to think that p because we do not have a[n unconscious] belief that not-p.").

140. The state of unconsciousness intended here is that of modest unawareness, not deep unawareness. *See supra* notes 51–54 and accompanying text (describing the distinction between modest and deep unawareness).

141. *See supra* notes 126–29 and accompanying text (describing some of these mechanisms). Desire has also been said to be able to trigger so-called "cold" or unmotivated cognitive biases or heuristics, such as the "confirmation bias" and the "availability heuristic," *see* MELE, *supra* note 131, at 28–29, with respect to which legal scholars have recently devoted considerable attention. The standard citation in the psychological literature regarding such biases and heuristics is RICHARD NISBETT & LEE ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT (1980). For a more recent account of the relevant literature, see HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT (Thomas Gilovich et al. eds., 2002).

142. MILO, *supra* note 89, at 114. *See also* Bach, *Thinking and Believing*, *supra* note 129, at 105 ("In self-deception, unlike blindness or denial, the truth is dangerously close at hand."); Scott-Kakures, *supra* note 132, at 32 ("An agent who is self-deceived somehow *knows* better.").

143. This account of self-deception is prominently associated with the work of Robert Audi. *See, e.g.*, AUDI, *supra* note 136; Robert Audi, *Self-Deception, Rationalization, and Reasons for Acting*, *in* PERSPECTIVES ON SELF-DECEPTION, *supra* note 131, at 92; Robert Audi, *Self-Deception and Rationality*, *in* SELF-DECEPTION AND SELF-UNDERSTANDING, *supra* note 131, at 169; Robert Audi, *Epistemic Disavowals and Self-Deception*, 57 PERSONALIST 378 (1976); Robert Audi, *Self-Deception and Practical Reasoning*, 19 CAN. J. PHIL. 246 (1989). Others have advanced what seem to be materially similar accounts. *See* COHEN, *supra* note 108, at 142 ("[The] self-deceiver . . . consciously accepts that p, [but he] does not *consciously* believe that not-p."); Bach, *supra* note 126, at 364–65 (claiming that the self-deceived actor believes that p but "avoid[s] the sustained or recurrent thought of p"); José Luis Bermúdez, *Defending Intentionalist Accounts of Self-Deception*, 20 BEHAV. & BRAIN SCI. 107, 107 (1997) (describing the self-deceived actor as one who "sincerely affirm[s] p and yet [has] an inferentially insulated belief that not-p"); Herbert Fingarette, *Self-Deception Needs No Explaining*, 48 PHIL. Q. 289, 296 (1998) (arguing that the self-deceiver "secretly" knows the truth "in [his] heart" or "deep down" but that the truth "is, in a crucial way, a secret even from [him]self").

This tension in turn gives him an internal reason to exercise doxastic self-control.

Self-deception so conceived can be understood as the mirror image of nonwillful ignorance. In cases of nonwillful ignorance the actor's desire not to believe that p^* prevents (through various mechanisms) him from forming the belief that p^*, although it does not succeed in preventing him from suspecting that p^*. Doxastic self-control is needed to check desire's influence so as to allow the belief to form, lest desire eventually prevail and cause the actor's suspicion itself to evaporate. In contrast, in cases of self-deception the actor's desire not to believe that p^* does not prevent the formation of the belief that p^*, but it does prevent (again through various mechanisms) that belief from rising to the requisite level of awareness. Doxastic self-control is needed to allow the unconscious belief to come to the surface, lest desire eventually prevail and push it ever deeper into the unconscious.

2. The Conative Condition.—All else being equal, an actor who, due to the causal influence of desire, fails to form the conscious belief that p can fairly be expected to exercise doxastic self-control so as to form that belief, provided he either suspects that p or unconsciously believes that p. In addition to this cognitive condition, however, a conative condition must also be met. For an actor's failure to control some ignorance-causing desires is culpable, but his failure to control other ignorance-causing desires is excusable. The best way to explain this condition is to examine the defense of duress.[144]

Duress is commonly portrayed as an excuse based on the principle that an actor should be excused if and because he lacked a "fair opportunity" to conform his conduct to the law's demands.[145] An actor who acts under duress chooses to do that which the law prohibits. He could have chosen instead to conform his conduct to the law's demands. He had the capacity so to conform, but he chose not to exercise that capacity. The law nonetheless recognizes that under certain circumstances it would be unfair to expect an

144. The discussion in the text focuses on the *content* or *object* of the ignorance-causing desire. If the strength of the desire, whatever its content or object, can fairly be characterized as compulsive or irresistible, then the actor should be excused for failing to control its influence on his beliefs, just as an actor should be excused for failing to control the influence of such a desire on his action. *But cf.* 2 PAUL H. ROBINSON, CRIMINAL LAW DEFENSES § 173(e)(1), at 83 (Supp. 2006–2007) ("[A] number of jurisdictions . . . have either declined to adopt an insanity defense that provides an excuse on the basis of control impairment or . . . have revised their insanity defense to eliminate an excuse based on control impairment.").

145. *See* Joshua Dressler, *Exegesis of the Law of Duress: Justifying the Excuse and Searching for Its Proper Limits*, 62 S. CAL. L. REV. 1331, 1365 (1989) ("A person acting under duress is excused, although he possessed the capacity to make the right choice, if he lacked a fair opportunity to act lawfully or, slightly more accurately, if he lacked a fair opportunity to avoid acting unlawfully."). In rare instances duress may also function as an incapacity excuse. *See* R.A. Duff, *Rule-Violations and Wrongdoings*, *in* CRIMINAL LAW THEORY, *supra* note 47, at 47, 64 ("[In some cases of duress, we] wish, and he wishes, that he could have resisted [the threat] (giving in was not justified): but he could not—his 'will' was 'overborne.'").

actor to choose conformity over disobedience. The law recognizes that under some circumstances it cannot fairly insist that an actor live up to the demands it imposes. Accordingly, under these circumstances the law offers duress as an excuse.

Duress is commonly portrayed as an excuse, but it might better be portrayed as a special form of justification or permission.[146] On this view, the law's own norms of permissible conduct do not allow the actor to act as he does, despite the duress under which he labors. The law expects him to conform. At the same time, however, the law condescends or defers to some other set of norms according to which the actor's conduct is permissible. Whatever the source and content of these norms—agent-relative moral norms,[147] role norms,[148] social norms[149]—they permit the actor to act as he

146. *See, e.g.*, MOORE, *supra* note 62, at 561 (arguing that the unfair opportunity branch of the choice theory of excuse "could be called the 'failed justification' idea of excuse"); ALAN WERTHEIMER, COERCION 168 (1987) (explaining duress as an "agent-relative" justification); Larry Alexander, *A Unified Excuse of Preemptive Self-Protection*, 74 NOTRE DAME L. REV. 1475, 1495 (1999) (explaining duress as a "personal" justification "reflecting the moral permissibility of a defendant's giving more weight in the moral calculus to his and his family's interests than those interests would be given from an impersonal perspective"); Claire O. Finkelstein, *Duress: A Philosophical Account of the Defense in Law*, 37 ARIZ. L. REV. 251, 280 (1995) [hereinafter Finkelstein, *Duress*] (duress as agent-relative justification); Claire Finkelstein, *Excuses and Dispositions in Criminal Law*, 6 BUFF. CRIM. L. REV. 317, 320 (2002) (arguing that "rational excuses" involve "cases where the agent acts for personal reasons, rather than for . . . impersonal considerations of social welfare"); Dan M. Kahan & Martha C. Nussbaum, *Two Conceptions of Emotion in Criminal Law*, 96 COLUM. L. REV. 269, 337 (1996) (duress as agent-relative justification); Robert F. Schopp, *Justification Defenses and Just Convictions*, 24 PAC. L.J. 1233, 1311 (1993) (duress as justification under "conventional social morality"); *see also* Craig L. Carr, *Duress and Criminal Responsibility*, 10 LAW & PHIL. 161, 183 (1991) (arguing that duress provides a defense in "those extraordinary cases that involve disobedience to law in response to a moral dilemma"); Dressler, *supra* note 145, at 1354–55 (arguing that an agent-relative theory of duress is "plausible[, but that even] if it is accepted, it would have rather limited applicability"); John Gardner, *The Gist of Excuses*, 1 BUFF. CRIM. L. REV. 575, 597–98 (1998) (arguing that the gist of an excuse, including that of duress, is that one has "lived up to" the applicable standards of character based on some conception of appropriate role); Stephen J. Morse, *Excusing and the New Excuse Defenses: A Legal and Conceptual Review*, 23 CRIME & JUST. 329, 341 (1998) ("Deciding which choices are too hard, that is, which threats might cause a person of reasonable firmness to yield and to do wrong [and thus be excused on the basis of duress], is of course a *normative* matter." (emphasis added)). *But see* Peter Westen & James Mangiafico, *The Criminal Defense of Duress: A Justification, Not an Excuse—And Why It Matters*, 6 BUFF. CRIM. L. REV. 833, 914 (2003) (claiming that an actor who validly claims duress is justified in the eyes of the law because he has "conformed to what the *law* ultimately regards as minimally acceptable conduct" (emphasis added)). For a brief, but in my view, effective critique of the Westen–Mangiafico thesis, see Mitchell N. Berman, *Justification and Excuse, Law and Morality*, 53 DUKE L.J. 1, 72–73 (2003).

147. *See, e.g.*, Finkelstein, *Duress*, *supra* note 146, at 280 ("[W]here the [actor's] reason for violating the [law's] prohibitory norm is *agent-relative*, that is, identified by its connection to the interests of the person whose reason it is, we demand only acceptance and understanding as a basis for exoneration."); Kahan & Nussbaum, *supra* note 146, at 337 (arguing that the norms governing the availability of duress are agent-relative).

148. *See, e.g.*, Gardner, *supra* note 146, at 593 (arguing that the availability of duress as a defense should "vary . . . according to the standards applicable to roles which the defendant occupies").

does and to satisfy the desire he seeks to satisfy in so acting, even though the law's own norms extend no such permission. So, for example, prevailing social norms might permit a woman to commit armed robbery in the face of a threat to injure her child if she refused, even if the law's own norms of permissible conduct would not.[150] In other words, prevailing social norms might permit the woman to satisfy her desire to save her child's life, even if the law's own norms would not.

Although the law commonly says when it will not entertain a claim of duress—when, for example, the actor kills to avoid the threatened harm[151] or when the threat comes from a source other than a human being[152]—it otherwise delegates to the trier of fact the task of judging when a claim of duress should prevail. For example, the Model Penal Code asks the trier of fact to decide whether a "person of reasonable firmness" in the actor's situation would have done as the actor did.[153] If the trier of fact decides that prevailing standards of reasonableness would have permitted the actor to act as he did, then punishing him for so acting would be unfair, and accordingly, he should be excused, even though he acted impermissibly according to the law's own norms of permissible conduct.

The law sometimes imposes demands that are too hard to satisfy. It sometimes asks us to control desires we cannot fairly or reasonably be expected to control. If an actor fails to control these desires and instead endorses and executes those desires in action, the law does not condone his failure. The law does not permit the woman to commit armed robbery, de-

149. *See, e.g.*, Dressler, *supra* note 145, at 1334 ("[An] actor should be excused [on grounds of duress] only if he attained or reflected *society's* legitimate expectations of moral strength." (emphasis added)); Kahan & Nussbaum, *supra* note 146, at 337 (arguing that the norms governing the availability of duress are "*social* norms that define the legitimate love of one's own" (emphasis added)).

150. *See* Kahan & Nussbaum, *supra* note 146, at 333 (describing such a case as one in which the actor "can assert duress").

151. *See, e.g.*, DRESSLER, *supra* note 56, at § 23.01[B], at 297.

152. *See id.* § 23.01[B], at 298. Although the Model Penal Code's formulation of the duress defense, like that of the common law, requires the threat to come from a human being, it rejects the common law's refusal to allow a defendant to claim duress to a charge of murder. *See* MODEL PENAL CODE § 2.09 (1962). For an even broader formulation than that of the Code, according to which an actor should be excused *whenever* he "committed [the crime] to avoid harm to himself or others, and a 'person of reasonable firmness' in the defendant's situation would have committed the crime," see Alexander, *supra* note 146, at 1494.

153. MODEL PENAL CODE § 2.09(1) (1962). To be more precise, the Model Penal Code formulation of duress asks whether a "person of reasonable firmness in [the actor's] situation would have been *unable* to resist." *Id.* (emphasis added). The Code therefore tries uncomfortably to combine in its formulation of the defenses duress-as-an-incapacity excuse and duress-as-a-no-fair-opportunity excuse, inasmuch as it makes the defense available only when the threat facing the actor would have rendered a person of reasonable firmness unable to resist. *Cf.* Dressler, *supra* note 145, at 1367 n.195 (noting that the Code "seems to treat duress as an incapacity-oriented excuse"). As a no-fair-opportunity excuse, duress would be available if a reasonable person in the actor's situation would not have resisted in the face of the threat directed toward the actor. As an incapacity excuse, duress would be available if the actor was unable to resist in the face of the threat directed toward him.

spite the threat to her child. It continues to insist that her conduct was impermissible and that she should have exercised control over the desire prompting her to act, even if under prevailing norms of reasonableness the exercise of such self-control would have been saintly or supererogatory.[154] At the same time, the law itself recognizes that its demands are sometimes too much to bear, and when they are, that it should excuse, but not condone, an actor's failure to bear them.

Duress excuses an actor for acting contrary to the law's demands, but the logic of duress extends beyond action to belief. An actor should be excused for failing to exercise doxastic self-control, such that he ends up failing to form the conscious belief that *p* when the available evidence otherwise supports the formation of that belief, if and because demanding such self-control would, as in the case of an action chosen under duress, be an unfair demand to make. In other words, although the law itself expects us to control any and all desires causing us to fail to form the conscious belief that *p* when the available evidence supports the formation of that belief, it should nonetheless recognize that failing to control the influence of some desires on belief constitutes a failure it ought to excuse, just as it recognizes through the doctrine of duress that failing to control the influence of some desires on action constitutes a failure it ought to excuse.

Which desires the law should excuse an actor's failure to control will depend on the norms to which it defers, and when it is prepared to defer to them. When the identity or application of those norms is indeterminate or controversial, the availability of duress will likewise be indeterminate or controversial. Consider again the case of Walter and Bernice Williams. Their claim to the contrary notwithstanding, assume they did believe their son's life would be in danger if they did nothing but give him aspirin. If the reason they did nothing but give him aspirin was their desire not to risk losing him to the welfare authorities, was their failure to act, though impermissible, nonetheless excusable? Would prevailing norms of reasonableness have permitted the Williamses to do nothing under the circumstances, other than give aspirin?

The welfare authorities were, at least so far as the Williamses were concerned, issuing a threat: Don't take your child to see a doctor, because if you do, we might take your child away from you. But should a parent ever be allowed to risk a child's life in order to avoid a risk of losing the child? What if the Williamses believed the risk of death was very small but the risk of losing him almost certain? Would that make a difference? Does the state,

154. *See, e.g.*, Duff, *supra* note 145, at 66 ("[W]hat citizens should ideally do is more than we, or the law, can properly *demand* that they do, on pain of condemnation if they do not; and that is why, whilst we do not regard such an agent's action as justified, we excuse her."); Berman, *supra* note 146, at 73 ("[T]he law could conceive of its own substantive, action-directing norms as sometimes demanding something closer to moral heroism."); Dressler, *supra* note 145, at 1367 (implying that the law sometimes demands "[]saintly moral strength" (emphasis omitted)).

being the source of the perceived threat, lose its standing to punish? If the answers to these questions prompt disagreement, as they doubtless will, the source of that disagreement resides in the indeterminacy of the norms governing when we are prepared to say that the choice facing an actor is or is not unfair.

The same indeterminacy spills over into the real case. The Williamses claimed not to have believed their inaction placed their son's life at risk. If they nonetheless suspected they were placing his life at risk, or if they unconsciously believed they were placing his life at risk, it would not, all else being equal, have been unfair to expect them to control the desire blinding them to the truth. But all else is not equal. For if their failure to exercise self-control, and thus their failure to realize the risk, was due to their desire not to chance losing their child, would it be unfair to expect a parent to control such a desire? Here is where the indeterminacy enters. If it would not have been unfair to expect such control, then the Williamses were rightly convicted of involuntary manslaughter. But if it would have been unfair, then although they could have controlled that desire, and although the law's own norms would have demanded such control, their failure to do so would nonetheless be a failure the law should excuse.

The case of Sam and Tiffany is not nearly so difficult. In their case the desire causing them to be blind to the lethal risk facing their child—the desire to climb higher on the social ladder—was stronger than it should have been. Indeed, we know it was stronger than it should have been precisely because it prevented them from seeing the risk to their son's life. Consequently, assuming they either suspected their child's life was in danger or unconsciously believed it was in danger, then they can fairly be punished for failing to control that desire. They are not punished for some prior culpable choice they may or may not have made. Nor are they punished for possessing a blameworthy desire blinding them to the truth. Instead, they are punished for failing to control that desire when they could and fairly should have controlled it.

IV. Conclusion

When the law legitimately imposes retributive punishment on an actor for involuntary manslaughter, it does so not because he made some prior culpable choice causing him later to be ignorant. Nor does it punish him for some hypothetical choice he might have made if, contrary to fact, he had believed he was risking another's life. Instead, it punishes him if and because he failed to exercise doxastic self-control over the influence of his desires on his belief when he could and fairly should have exercised such self-control.

An actor's desires sometimes make it impossible for him then and there to form the conscious belief that he is risking another's life. Yet insofar as he retains the capacity to alter the balance of his competing desires, he is obliged to exercise that capacity, thereby permitting the risk confronting him

to come clearly into focus, provided he either suspects the risk exists or else unconsciously believes it exists, and provided too that the desire blinding him to the risk is a desire he should have controlled. If these conditions obtain, and the actor nonetheless fails to exercise such self-control, he is a fair candidate for retributive punishment. He not only could have controlled the ignorance-causing desire, but demanding the exercise of that capacity would be a fair demand to make.

When an actor freely and wholeheartedly chooses to do that which he knows the law demands he not do, he can be said willfully to *defy* the law. He willfully disobeys. He symbolically sets himself above the law. The immediate aim of retributive punishment directed at such an offender is to humble his defiant will, although ideally the offender will come to embrace his punishment as part of the price he must pay to make amends for his defiance. Having paid that price he will have earned forgiveness and the right to rejoin the community from whom he has separated himself through the commission of the offense.[155] He will have atoned for his crime.

An actor guilty of involuntary manslaughter does not defy the law. He cannot defy it because he does not see the risk he is taking, and if he does not see the risk he is taking, he cannot know that the risk is one the law demands he not take. The hardship imposed on such an actor is thus imposed, not to punish him for defiance, but rather to censure him for weakness, for failing to control the desire causing his blindness when he could have exercised such control, and when the demand for its exercise was fair to make. It should also, and perhaps more importantly, teach him how better to exercise his capacity for doxastic self-control and so become less vulnerable to the power of ignorance-causing desire.[156]

155. For a more detailed explanation of the ideas contained in this paragraph, see generally Stephen P. Garvey, *Punishment as Atonement*, 46 UCLA L. REV. 1801 (1999).

156. *Cf.* Robert Merrihew Adams, *The Virtue of Faith*, 3 FAITH & PHIL. 3, 5 (1984) ("The purpose of identifying cognitive failures as sins is not to find a stick to beat the sinner, but rather to learn of what we have to repent of.").