ONLY GUILTY MINDS

By Michael Serota

This conference submission is a draft excerpt from a larger project with the tentative title “Only Guilty Minds.” That project makes the case for viewing mens rea reform, broadly construed, as a meaningful component of a modern criminal justice reform agenda. In so doing, I will argue that the primary beneficiaries of mens rea reform—or at least the version of mens rea reform developed in my paper—are no different than those who disparately suffer from every other problem in the criminal justice system: the poor, the underserved, and people of color. For purposes of this conference, however, I’m only submitting one component of this project, namely, its theoretical foundation. That foundation, which constitutes Part I of this project, provides an introduction to mens rea, and is organized around three main questions: (1) What is mens rea?; (2) Why do we care about mens rea?; and (3) Why should we care about mens rea?

There are two goals motivating this submission. The first is to offer what I hope is a clear and accessible framing of a topic that has been a source of confusion among judges, lawyers, and legislators. This is not the only framing of mens rea possible, but it is the framing that I believe is most intuitive and which may best serve the goals of the conference. The second is for the purpose of receiving feedback. Having worked on mens rea policy as a legislative drafter for over six years, this project is my preliminary attempt to express what I hope is a broadly appealing vision of the law of guilty minds that is rooted in our everyday judgments of blame and moral responsibility. Through this effort, I’m seeking to develop an understanding of mens rea that synthesizes, refines, and builds upon many of the most important scholarly contributions on the topic in a way that is persuasive not only to criminal law scholars but also to the policymakers on whom criminal justice reform depends, as well as to their constituents who stand to benefit from it.

Please note: readers should feel free to ignore the footnotes, which are quite extensive even by academic standards, and more than double the length of the 12,000 words of main text. I opted to include them because they provide information that those not well-versed in the field may find useful or of interest in learning more about the topic at the heart of this conference.
Part I of Only Guilty Minds

MENS REA & THE GUILTY MIND—CONCEPTS, PRINCIPLES, AND POLICIES

A. What is Mens Rea?

The topic of mens rea is a central preoccupation of American criminal law. Indeed, from criminal codes, to judicial opinions, to scholarship, legal writing on mens rea is pervasive. And those who took substantive criminal law at an American law school likely spent much of their time reading this writing in a class for which mens rea is the primary focus. But for a topic that has received so much attention from the American legal profession, it is striking just how little agreement there is as to what mens rea actually is.

Roughly translated, the term mens rea means “guilty mind” in Latin. The significance of this term is derived from the oft-cited Latin phrase “actus non facit reum nisi mens rea,” which

---

1 See, e.g., Douglas Husak, "Broad" Culpability and the Retributivist Dream, 9 OHIO ST. J. CRIM. L. 449, 449 (2012) (“The criminal law is centrally concerned with culpability or mens rea—roughly, the mental (or quasi-mental) components of blame.”); Gideon Yaffe, The Point of Mens Rea: The Case of Willful Ignorance, 12 CRIM. L. & PHILO. 19, 25 (2018) (“Criminal law exhibits a well-known fixation with the defendant’s mind, a fixation that we do not find in other areas of law, including areas in which the mental states of the parties matter to liability.”); Gary V. Dubin, Mens Rea Reconsidered: A Plea for a Due Process Concept of Criminal Responsibility, 18 STAN. L. REV. 322, 325 (1966) (observing the “nearly deified legal status” of mens rea).

2 MODEL PENAL CODE & COMMENTARIES (AM. LAW INST. 1985).


6 See, e.g., Stephen P. Garvey, Authority, Ignorance, and the Guilty Mind, 67 SMU L. REV. 545, 545–46 (2014) (“The problem is that no one really knows, or if they know, haven't clearly said, in what the guilty mind consists.”); Francis Bowes Sayre, The Present Signification of Mens Rea in the Criminal Law, in HARVARD LEGAL ESSAYS 399, 402 (1934) (“The term mens rea is “chameleon-like” in nature.”); FLETCHER, supra note 4, at 398 (“[T]here is no term fraught with greater ambiguity than that venerable Latin phrase that haunts the Anglo-American criminal law: mens rea.”); Sanford Kadish, The Decline of Innocence, 26 CAMBRIDGE L.J. 273, 274–75 (1968) (“[T]he term ‘mens rea’ is rivaled [by few legal terms] for the varieties of senses in which it has been used and for the quantity of obfuscation it has created.”); Sayre, supra note 4, at 974 (“[W]hen it comes to attaching a precise meaning to mens rea, courts and writers are in hopeless disagreement.”).

7 BLACK’S LAW DICTIONARY, 1137.
means “an act is not culpable unless the mind is guilty.” 8 In America, we have our own way of reciting this axiom, which the U.S. Supreme Court first articulated in Morissette v. United States (1952) and American legal culture has fervently recounted ever since: “Crime, as a compound concept, generally [requires the] concurrence of an evil-meaning mind with an evil-doing hand.” 9

The basic idea captured by these aphorisms is intuitive, and it plays a central role in the criminal law. But prior to discussing them, it’s worth briefly introducing a few concepts that are important to any discussion of mens rea but also are a perennial source of confusion. 10

The most central concept is blame, which is—among many other things 11—a judgment of disapproval of a human action or quality, 12 and a social practice that is as familiar to most of us as

8 Garvey, supra note 6, at 545 (observing that “[t]he phrase is usually traced to St. Augustine's Sermon 180 on perjury.”) (citing Saint Augustine, Sermon 180, in 5 THE WORKS OF SAINT AUGUSTINE: A TRANSLATION FOR THE 21ST CENTURY—SERMONS, 314, 315 (JOHN E. ROTELLE, O.S.A. ED., EDMUND HILL, O.P. TRANS., 1992)); see also 4 WILLIAM BLACKSTONE, COMMENTARIES *21 (noting that “an unwarrantable act without a vicious will is no crime at all”).
10 The concepts discussed below—blame, responsibility, wrongfulness, and culpability—arise at the intersection of moral psychology and criminal law theory, for which they play a central (if at times slightly different) role. David O. Brink, The Nature and Significance of Culpability, 13 CRIM. L. & PHIL. 347, 347–48 (2019) (“Even if these shared concepts play somewhat different roles in these two domains, there is clearly significant overlap in their nature and demands. In fact, given the importance of moral ideas to the formation and reform of criminal law principles and practices and the effect of well-settled criminal law doctrine on our moral assumptions and beliefs, we should expect mutual influence and interaction between these domains.”).
11 Here I only endeavor to provide a simple, but useful, working understanding of blame that serves the purpose of this Part. In reality, however, the concept of blame is just as capacious—and perhaps even more inescapable than—that of mens rea. David Shoemaker, Blameworthy but Unblamable: A Paradox of Corporate Responsibility, 17 GEO. J.L. & PUB. POL’Y 897, 905 (2019) (“Blame is incredibly capacious, likely uncatchable by a single theoretical account[.]”). This capaciousness is reflected in the fast-growing philosophical literature on the nature of blame, which has produced a number of understandings—including the following views:

(1) Blame is a reactive attitude, such as anger, resentment, and indignation. P.F. Strawson, Freedom and Resentment in FREE WILL, 72-93 (2d ed. 2003)); see R. JAY WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS, 75 (1994) (it would be “strange to suppose that one might blame another person without feeling an attitude of indignation or resentment toward the person…”).
(2) Blame is a pair of mental activities, namely, (1) a judgment that someone has done something reflecting “attitudes toward others that impairs the relations that others can have with him or her”; and (2) a modification of one’s own attitudes, intentions, or dispositions in a way appropriate to that relationship impairment. THOMAS SCANLON, MORAL DIMENSIONS: PERMISSIBILITY, MEANING, BLAME, 127-28 (2008).
(3) Blame is an attitudinal or behavioral response, which repudiates or takes a stand against someone’s immoral treatment, and is often accompanied by protest that, inter alia, aims to get the offender and others in the moral community to acknowledge what the offender did. Angela Smith, Moral Blame & Moral Protest, in BLAME: ITS NATURE AND NORMS 41-47 (ed. Justin Coates and Neal Tognazinni, 2013).
(4) Blame signals the blamer’s commitment to a set of norms. David Shoemaker & Manuel Vargas, Moral Torch Fishing: A Signaling Theory of Blame, Nous (2019).

For a recent edited volume that provides a good overview of the field, see Justin Coates and Neal Tognazinni, BLAME: ITS NATURE AND NORMS (2013).
12 Sanford H. Kadish, Complicity, Cause and Blame: A Study in the Interpretation of Doctrine, 73 CAL. L. REV. 323, 329 (1985) (“Blame and its correlative, praise, serve as expressions of our disapproval or approval of some human action or quality.”); id. (“[Blame] is not intrinsically an expressive action, but a judgment of disapproval. It is an internal evaluation that need not be expressed. Blame is the sentiment of disapproval itself.”).
the air we breathe.\textsuperscript{13} Most often, we blame people for the (bad) things they do, which brings with it a judgment of responsibility for the negative effects caused or threatened by the offensive conduct.\textsuperscript{14} Generally speaking, our blaming and responsibility practices rest upon two main evaluative criteria: wrongfulness and culpability.

Wrongfulness is typically understood to be an attribute of \textit{acts},\textsuperscript{15} whereas culpability is an attribute of \textit{actors}.\textsuperscript{16} A person’s act is wrongful when (and to the extent) it violates an accepted norm of conduct.\textsuperscript{17} Many categories of these behavioral norms exist, ranging from the aesthetic (e.g., “color within the lines”) to the professional (e.g., “write criminal law papers clearly”).\textsuperscript{18} But perhaps the most important category—or at least the category that is most central to our blaming and responsibility practices—is that of moral norms, which are norms that implicate the well-being or interests of other people (e.g., “don’t force people to read unclear criminal law papers”).\textsuperscript{19}

Conduct that violates a moral norm is morally wrongful.\textsuperscript{20} And a finding of morally wrongful conduct is critical to our expressions of blame and responsibility.\textsuperscript{21} But it’s not everything. Because our blame and responsibility judgements also hinge upon a finding of culpability, which focuses on the mind of the person, in contrast to the thing that he or she did.\textsuperscript{22} We can say, for these introductory purposes, that culpability is comprised of those mental characteristics that enable us to attribute wrongful conduct to the actor in a way that makes moral blame and responsibility appropriate\textsuperscript{23}—or what this paper will often refer to as a “guilty mind.”

\begin{itemize}
  \item \textsuperscript{13} \textit{Id.} at 329 (“Attributing blame is a pervasive human phenomenon. It is one way in which we order and make sense of social experience and it is reflected in our language and social practices.”).
  \item \textsuperscript{14} \textit{Id.} (“[B]lame entails a judgment of responsibility . . . One who presents a poor appearance is not blamed for it unless he is responsible for it, as by poor judgment or lack of care in dress or grooming. The runner who loses a race may be blamed if he failed to train properly, but not if he did the best he could.”); \textit{see, e.g.}, Meir Dan-Cohen, \textit{Responsibility and the Boundaries of the Self}, 105 HARV. L. REV. 959, 959–60 (1992) (“[Blame involves] ascribing moral responsibility for the negative effects of one’s behavior”).
  \item \textsuperscript{15} \textit{See, e.g.}, YAFFE, supra note 4, at 66-68 (“Wrongfulness is a property of actions. An action is wrongful (or, equivalently, an instance of wrongdoing) just in case it is in violation of a norm; it is not-to-be-done.”); KADISH, supra note 12, at 329-32.
  \item \textsuperscript{16} \textit{See, e.g.}, YAFFE, supra note 4, at 66–68 (“Culpability is not, like wrongfulness, a property of acts, but, instead, a property of agents. However, agents are not culpable, full stop. Rather, they are culpable for acts, and not act types, but act tokens.”); KADISH, supra note 12, at 329-32.
  \item \textsuperscript{17} \textit{See, e.g.}, YAFFE, supra note 4, at 66–68 (“An action is wrongful (or, equivalently, an instance of wrongdoing) just in case it is in violation of a norm; it is not-to-be-done.”); KADISH, supra note 12, at 329-32.
  \item \textsuperscript{18} \textit{See, e.g.}, YAFFE, supra note 4, at 66–68 (“[There are different] sets of norms. If you’re painting a portrait, it might be wrong to paint the eyes blue. It is not morally wrong, but aesthetically wrong, or maybe wrong because it violates norms of a particular genre of painting, such as realistic portraiture.”); KADISH, \textit{supra} note 12, at 329-32.
  \item \textsuperscript{19} \textit{See, e.g.}, YAFFE, supra note 4, at 66-68 (“Typically, the relevant kind of wrongfulness is moral wrongfulness and so the relevant set of norms are moral norms.”); KADISH, \textit{supra} note 12, at 329-32.
  \item \textsuperscript{20} \textit{See, e.g.}, YAFFE, \textit{supra} note 4, at 66-68; KADISH, \textit{supra} note 12, at 329-32
  \item \textsuperscript{21} \textit{See, e.g} YAFFE, \textit{supra} note 4, at 68 (“To say that someone is “culpable for an act” is to imply two things: he is responsible for the act, and the act is wrongful. So culpability is a species of responsibility: it is responsibility for wrongful action.”); KADISH, \textit{supra} note 12, at 329-32.
  \item \textsuperscript{22} \textit{See, e.g.}, YAFFE, \textit{supra} note 4, at 68; KADISH, \textit{supra} note 12, at 329-32.
  \item \textsuperscript{23} \textit{See, e.g.}, KADISH,\textit{supra} note 4; Peter Westen, \textit{Individualizing the Reasonable Person in Criminal Law}, 2 CRIM. L. & PHIL. 137 (2008); Husak, \textit{supra} note 1. It is for this reason, as Thomas Scanlon writes, that:
  \begin{quote}
    [t]he blameworthiness of an action depends, in ways that wrongness generally does not, on the reasons for which a person acted and the conditions under which he or she did so. So it can be appropriate to say such things as, “Yes, what she did was certainly wrong, but you shouldn’t blame
  \end{quote}
I will have much more say about culpability and the guilty mind throughout this part. But for now, let us return to the mens rea aphorisms mentioned at the beginning of this section. At the heart of both of them is an intuitive idea: we surely regret it when people do the wrong thing, but the criminal justice system should only blame people—through its formalized methods of conviction and punishment—for violating the moral norms proscribed by the criminal law when they do so with a guilty mind. To illustrate this principle of mens rea, consider a few simple examples:

(1) A confused restaurant-goer takes someone else’s umbrella on her way out, having reasonably mistaken it for her own.

(2) A conscientious klutz accidentally drops her books on her best friend’s foot, notwithstanding her best efforts at being careful.

(3) An emotionally distraught driver accidentally crashes into a parked car, having been distracted by news that her only child has been diagnosed with terminal cancer.

Have these individuals engaged in morally wrongful conduct? No question—morality generally instructs us not to steal, injure, or destroy the property of others. Might we expect these individuals to civilly compensate the victims for their loss or injuries? Perhaps. But do we perceive their behavior as sufficiently blameworthy to merit punishment? Unlikely. And why is that? Because of the psychological phenomena at issue. No doubt, the criminal law codifies moral norms against taking what’s not yours, harming other people, or destroying other people’s property, in the absence of adequate justification. Nevertheless, the mental phenomena accompanying these prohibited acts—for example, a reasonable mistake in the first hypothetical, and the unavoidable accidents in the second and third—are insufficiently culpable to justify holding the person criminally responsible for them.

But if that’s true, then it’s reasonable to ask: what kind of mental state would be sufficiently culpable? One intuitively obvious answer is a culpable intent—consider, for example, how the situation would differ if the above individuals had intended to steal, injure, or destroy. Another compelling response is culpable awareness—imagine, for example, if the above individuals had been aware that the taking, injury, or destruction would, or at least might, have resulted from their conduct. A third possibility, which is perhaps slightly less obvious though still intuitive for many, is culpable inadvertence—that is, a failure to be aware of the relevant risks to the interests of others under circumstances in which a reasonable person, if placed in the same situation, would have perceived them.

Scanlon, supra note 10. I will have more to say about the ideas expressed in this passage later on in the paper.

24 A moral norm codified by the criminal law is a legal norm. But not all legal norms are moral norms—such as, for example, the legal norm against jaywalking. See, e.g., Yaffe, supra note 4, at 67. Nevertheless, for purposes of this paper, I will largely limit my discussion to the moral norms codified by the criminal law.

25 See, e.g., Kadish, supra note 4, at 260–61 (1987); Westen, supra note 23, at 151.

26 See infra note 42 for discussion of the concept of justification and how it can be differentiated from that of excuse.
As it turns out, these mental states are precisely what the criminal law requires the
government to prove—typically, beyond a reasonable doubt—by way of the mens rea requirement
applicable to most statutes (and prosecutions).27 This is reflected in the standard culpable mental
hierarchy employed by a majority jurisdictions, which is comprised of four concepts—what most
jurisdictions refer to as purpose, knowledge, recklessness, and negligence.28 Here’s a quick and
simplified overview of what the terms that comprise this modern PKRN framework are understood
to mean:

(1) *Purpose* typically entails proof of a conscious desire to engage in wrongdoing29—i.e.,
to cause harm or act under prohibited circumstances.30 It is reflected where (for
example) X pulls the trigger of a loaded gun with the goal of killing V. If X is
successful, then the law says that X purposely killed V.

(2) *Knowledge* typically entails proof of a high level of awareness of wrongdoing31—i.e.,
that one’s conduct would result in harm or occur under prohibited circumstances.32 It
is reflected where (for example) X, a child rights advocate, blows up a manufacturing
facility that relies on youth labor, aware that V, the on-duty night guard who is
unaffiliated with the business, will almost surely die in the blast. If V does die in the
blast, then the law says that X knowingly killed V.

(3) *Recklessness* typically entails (among other requirements) proof of a low level of
awareness of wrongdoing33—i.e., that one’s conduct might result in harm or occur
under prohibited circumstances. It is reflected where (for example) X speeds through a
red light aware that it is substantially possible that his car will fatally hit V, a pedestrian
stepping into the crosswalk. If V is fatally struck, then the law says that X recklessly
killed V.34

27 “Most” is decidedly not “every,” however. *See infra* Part II (not included in conference submission).
28 The genesis and development of this hierarchy will be discussed *infra*, Part II (not included in conference
submission).
29 *See Model Penal Code § 2.02(2)(a) (AM. LAW INST., Proposed Official Draft 1962).*
30 This conscious desire to engage in wrongdoing is to be distinguished from wrongful conduct that is *voluntarily*
The voluntariness requirement, which implicates what is sometimes referred to as a “*present conduct intention,*” can
be “satisfied simply by showing that the actor did in fact intend to perform the bodily movements that he performed.”
The mere fact that someone engages in wrongdoing voluntarily is entirely consistent with strict liability. *See, e.g.*, *Buchanan v. United States*, 32 A.3d 990, 1000 (D.C. 2011) (Ruiz, J., concurring) (noting that the “intent to act”
interpretation of simple assault, if taken literally, would allow “the prosecution of individuals for criminal assault for
actions taken with a complete lack of culpability”).
31 *See Model Penal Code § 2.02(2)(b).*
32 Here, and with each of the following culpable mental states, I don’t mean awareness that one’s conduct might be
illegal—although that too would surely seem to implicate the blameworthiness of a person’s state of mind. *See, e.g.*, *Douglas Husak, Ignorance of Law: A Philosophical Inquiry* (2016). This Part largely omits discussion of
ignorance and mistakes of law for purposes of accessibility and simplicity, not because such issues aren’t important.
*See infra* note 89.
33 *See Model Penal Code § 2.02(2)(c).*
34 That is, provided the defendant’s conduct also satisfies the gross deviation standard governing recklessness
inquiries. *See id.*
Negligence typically entails proof of (among other requirements) an unreasonable failure to be aware of the risks associated with one’s conduct\textsuperscript{35}—i.e., failing to perceive that one’s conduct might result in harm or occur under prohibited circumstances where a reasonable person, if confronted with the same situation, would have perceived that risk. It is reflected where (for example) X speeds through a red light while reading his phone, unaware that it is substantially possible that his car will fatally hit V, a pedestrian stepping into the crosswalk. If V is fatally struck, then the law says that X negligently killed V.\textsuperscript{36}

Perhaps unsurprisingly, these four mental states are central to most discussions about mens rea—\textsuperscript{37}and this is equally true in theory,\textsuperscript{38} as a matter of legislative drafting,\textsuperscript{39} and in court.\textsuperscript{40} However, they are also far from the only mental states that matter to the criminal law. For example, it is entirely possible for someone to violate a criminal prohibition with a state of mind that satisfies the PKRN framework and yet still evade a conviction. That’s because the criminal law also recognizes the moral salience of a number of other mental characteristics that fall outside of this four-part framework yet still bear on a person’s blameworthiness. Consider, for example, each of the following situations:

(1) A mentally ill restaurant-goer \textit{purposely} takes someone else’s umbrella, but—due to the influence of her serious illness—is not able to appreciate that taking other people’s umbrellas is wrong.

(2) A hapless defender \textit{purposely} drops her books on the foot of someone she mistakenly perceives to be a robber about to pull out a firearm, only to discover that the alleged perpetrator is just her friend with a mediocre pandemic haircut.

(3) A coerced mother is asked to choose at gunpoint between \textit{purposely} slapping someone else’s child or her child being slapped by someone else—and after suffering an emotional breakdown, ultimately opts to do the slapping herself.

\textsuperscript{35}See id. § 2.02(2)(d).
\textsuperscript{36} That is, provided the defendant’s conduct also satisfies the gross deviation standard governing negligence inquiries. See id.
\textsuperscript{37} See, e.g., Garvey, supra note 8, at 546 (“Today’s criminal law relies on, and has no need for anything more than, \textit{mentes reae}, which are nothing more, and nothing more mysterious, than the various mental states that the state happens to have decided it must prove in order to convict an actor of this or that crime.”).
\textsuperscript{39} See, e.g., Robinson & Grall, supra note 38, at 692 (“Section 2.02 may appropriately be considered the representative modern American culpability scheme.”); D.C. Criminal Code Reform Commission, Report #2, First Draft Recommendations for Chapter 2 of the Revised Criminal Code - Basic Requirements of Offense Liability, https://ccrc.de.gov/node/1208152
No doubt, the purposeful mental states reflected in these situations\(^{41}\) would be sufficient to satisfy the government’s affirmative burden of proof in prosecutions for theft (in the first situation) or assault (in the latter two). However, the criminal law allows for the possibility that all three of these individuals should escape liability by virtue of other morally salient psychological phenomena—namely, the restaurant-goer’s substantially impaired moral judgment, the hapless defender’s virtuous motivation, and the coerced mother’s loss of control—which the criminal law

\(^{41}\) That is, the purpose to take someone else’s property, to injure by dropping, or to injure by slapping.
recognizes through affirmative defenses\textsuperscript{42} such as insanity,\textsuperscript{43} duress,\textsuperscript{44} and mistake as to a justification.\textsuperscript{45} That the criminal law would provide these \textit{excuses} is not surprising given the extent to which these phenomena seem to diminish blameworthiness. But are they mens rea?

\textsuperscript{42} There are two main categories of affirmative defenses: justifications and excuses. Both constitute a claim of personal innocence; however, each rest upon different grounds.

A justification defense rests upon the ground that the defendant’s conduct was not \textit{wrongful} and may even have been affirmatively desirable under the circumstances. \textit{Joshua Dressler, Understanding Criminal Law} § 16.03 (7th ed. 2019); (justified conduct is conduct that violates a criminal prohibition yet society considers it to be “a good thing, or the right or sensible thing, or a permissible thing to do” under the circumstances). As Sandy Kadish phrases it:

\begin{quote}
[The nature of the] claim here is that I did nothing wrong even though I violated the prohibition. This is so, I argue, because the crime’s definition of the forbidden conduct is incomplete. The law allows what the crime as defined prohibits where circumstances, specified elsewhere in the law, make my action the right thing to do: I killed, but only to keep an assailant from killing me or another; I entered another’s cabin in the wilderness, but only to obtain food and water to keep myself alive.
\end{quote}

Kadish, \textit{supra} note 4 at 258; \textit{see Model Penal Code} § 3.04(1) (defining the justification of self-defense); \textit{id.} at § 3.02(1)(a) (defining the justification of necessity).

An excuse defense, in contrast, rests upon the ground that defendant’s conduct, although wrongful, was not blameworthy under the circumstances. \textit{Dressler, supra} note 42, at § 16.04 (distinguishing justifications, which negate the harmfulness of an offense, with excuses, which negate its blameworthiness). In this way, a criminal defendant who asserts an excusing defense says, in essence, “I admit, or you have proved beyond a reasonable doubt, that I did something I should not have done, but I should not be held criminally accountable for my actions” due to a lack of culpability. \textit{Joshua Dressler, Reflections on Excusing Wrongdoers: Moral Theory, New Excuses and the Model Penal Code}, 19 \textit{Rutgers L.J.} 671, 675-76 (1988); \textit{see Kadish, supra} note 4, at 258.

\textsuperscript{43} Generally speaking, the excuse of insanity asserts the absence of blameworthiness due to defects of reasoning and moral judgment, which were the result of a mental disease and that precluded the person from complying with the law. Kadish, \textit{supra} note 4 at 261.

There are a number of approaches to insanity; however, the M’Naghten and Model Penal Code approaches are most common today. \textit{See} Paul H. Robinson et. al., \textit{The American Criminal Code: General Defenses}, 7 \textit{J. Legal Analysis} 37, 78 (2015) (providing a detailed overview of prevailing legal trends).

The M’Naghten test asks whether as a result of mental disease the defendant was unable to know the nature of his act, or that the act was wrong. \textit{See, e.g., Wayne R. Lafave, 1 Subst. Crim. L.} § 7.1 (3d. ed. 2018) (“[U]nder the prevailing M’Naghten rule (sometimes referred to as the right-wrong test) the defendant cannot be convicted if, at the time he committed the act, he was laboring under such a defect of reason, from a disease of the mind, as not to know the nature and quality of the act he was doing; or, if he did know it, as not to know he was doing what was wrong.”).

The Model Penal Code approach expands M’Naghten (among other ways) to ask as well whether a mental disease prevented the defendant from conforming his conduct to the requirements of the law. \textit{Model Penal Code} § 4.01(1) (“A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the criminality [wrongfulness] of his conduct or to conform his conduct to the requirements of law.”).

\begin{quote}
[T]here is,” as Joshua Dressler observes, “immense debate regarding the proper scope of the [insanity] defense”:
\end{quote}

Should the defense be limited to cognitive disabilities or include volitional incapacity? Should the criminal law revive the previously discredited product test of insanity? In jurisdictions applying the majority M’Naghten test, should they narrow the test to just one of the two prongs, and if so, which? Further, should the prong(s) be stated in terms of “knowledge” or the deeper concept of “appreciation” of right-from-wrong and/or the nature-and-quality of the defendant’s actions? Finally, regarding the right-from-wrong test, are we talking about legal or moral right-from-wrong?

Dressler, \textit{supra} note 42, at 256–57.
The criminal law is of two minds on this question.\textsuperscript{46} An older tradition takes the position that the latter set of issues qualifies as mens rea, even if the law treats them as affirmative defenses, on the view that mens rea is capacious enough to encompass any and all mental (or quasi-mental)

\begin{quote}
\textsuperscript{44} Generally speaking, the excuse of duress asserts the absence of blameworthiness due to a crime being committed under the command of another backed by such threats of physical injury that even a person of reasonable fortitude would have done the same. Kadish, \textit{supra} note 4, at 261.

There are two main approaches to duress. The first is the common law approach, which applies in a prosecution for an offense other than murder, where “the actor engaged in the proscribed conduct because he was coerced to do so by what he reasonably believed was an unlawful threat of imminent death or severe bodily injury to himself or another.” Paul H. Robinson et. al., \textit{supra} note 43, at 88. The second is the broader Model Penal Code approach, which applies to any offense that was committed because the defendant “was coerced to do so by the use of, or a threat to use, unlawful force against his person or the person of another, that a person of reasonable firmness in his situation would have been unable to resist.” \textit{MODEL PENAL CODE} § 2.09(1).

\textsuperscript{45} Generally speaking, the excuse of mistake as to a justification asserts the absence of blameworthiness because the person reasonably believed that the factual situation was such that his or her conduct was necessary and appropriate for any of the purposes that would establish a justification defense. Paul H. Robinson et. al., \textit{supra} note 43, at 75; see Paul H. Robinson, \textit{Crim. L. Def.} § 184 (mistake as to a justification). In the relevant situations, “A person’s conduct is objectively unjustified but the person subjectively, mistakenly believes that it is justified” (i.e., “the actor believes that her conduct avoids a greater harm, when in fact it does not”). Paul H. Robinson & John M. Darley, \textit{Testing Competing Theories of Justification}, 76 N.C. L. REV. 1095, 1101 (1998) (“An actor is excused for her conduct constituting an offense if her conduct would be justified had the attendant circumstances been as she believed them to be.”).

The most common version of this affirmative defense is mistaken self-defense, which excuses a non-aggressor who employs “a reasonable amount of force against his adversary when he reasonably believes (a) that he is in immediate danger of unlawful bodily harm from his adversary and (b) that the use of such force is necessary to avoid this danger.” \textit{LAFAVE}, \textit{supra} note 43, at 2 SUBST. CRIM. L. § 10.4 (“When [the defending party’s] belief is reasonable . . . he may be mistaken in his belief and still have the defense.”); compare \textit{MODEL PENAL CODE} § 3.04(1) (“Subject to the provisions of this Section and of Section 3.09, the use of force upon or toward another person is justifiable when the actor believes that such force is immediately necessary for the purpose of protecting himself against the use of unlawful force by such other person on the present occasion.”); see also Richard Singer, \textit{The Resurgence of Mens Rea: II—Honest but Unreasonable Mistake of Fact in Self Defense}, 28 B.C. L. REV. 459, 491 (1987).

It is important to acknowledge that the criminal law often formally refers to mistaken self-defense, and mistaken justifications more generally, as justifications instead of excuses. See, e.g., sources cited \textit{id}. However, this categorization is conceptually inconsistent with the intuitive beliefs that (1) justifications apply to acts which do in fact avoid a greater harm and (2) results matter to criminal liability (among others). See, e.g., Robinson & Darley, \textit{supra} note 45, at 1132 (summarizing lay and scholarly intuitions). For clarity’s sake, then, I will follow the more intuitive conception.


As Steve Garvey explains, this distinction has been framed in a number of ways—for example, he characterizes it as the choice between “mens rea and mentes reae,” while others describe it as the “distinction between culpability and elemental accounts of mens rea, or normative and descriptive accounts, or broad and narrow accounts.” Garvey, \textit{supra} note 6 at 546–47. I’m extremely reticent to add yet another framing of this distinction; however, I do believe that the different terminology employed in this article—“PKRN” and “guilty minds”—is more descriptively accurate, and therefore the most likely to ease the cognitive burden on readers who are not already familiarized with this debate.
aspects of blame.\textsuperscript{47} Let us call this the “guilty minds” view of mens rea.\textsuperscript{48} In contrast, a more contemporary tradition rejects this broader understanding, taking the position that mens rea is (or at least should only be used to refer to) the purpose, knowledge, recklessness, or negligence that comprises the government’s affirmative burden of proof.\textsuperscript{49} Let us call this the PKRN view of mens rea.\textsuperscript{50}

So, then, which perspective on mens rea is right—guilty minds or PKRN? Ultimately, I think the answer has to be both (or neither). That’s because the term mens rea serves two different functions in the criminal law, and each understanding is better situated to perform one of them. The first thing that the term mens rea does is serve as a moral concept, to be employed when thinking about the major policy choices that the criminal law makes (i.e., who should we punish and how severely). The second thing that the term mens rea does is serve as a legal construct, which enables us to segregate, articulate, and organize those policy choices. Arguably, the broader guilty minds understanding of mens rea is better suited for the first function, whereas the narrower PKRN version is better suited for the latter.

Understanding mens rea broadly, in terms of any and all aspects of a person’s mind that aggravate or diminish blame, is helpful for evaluating the big picture—that is, what makes a mind guilty, why should the criminal law care, and do our criminal laws adequately reflect this concern? But employing this understanding in the legal context—for example, when drafting criminal statutes or in judicially explicating them—is usually less helpful, and sometimes problematic. That’s because it effectively groups together different aspects of blameworthiness, which the criminal law (reasonably) deals with in procedurally different ways—for example, by treating PKRN as part of the government’s affirmative burden of proof, while treating insanity, duress, and (mistaken) self-defense as affirmative defenses to be raised by the accused.\textsuperscript{51}

\textsuperscript{47} For discussion of this older tradition, which views \textit{mens rea} as a “morally blameworthy state of mind,” along historical lines, see Dressler, supra note 42, at § 10.02. And for an outstanding effort to explore the conceptual boundaries of this broader view of \textit{mens rea}, see Husak, supra note 1, at 454–60.

\textsuperscript{48} Or what is also referred to as the “culpability,” “normative,” and “broader” accounts of mens rea. See Garvey, supra note 6, at 546.

\textsuperscript{49} This framing is rooted in the work of the Model Penal Code. See, e.g., \textit{Model Penal Code} § 2.02(1) (“Except as provided in Section 2.05, a person is not guilty of an offense unless he acted purposely, knowingly, recklessly or negligently, as the law may require, with respect to each material element of the offense.”); Herbert Wechsler, \textit{Codification of Criminal Law in the United States: The Model Penal Code}, 68 COLUM. L. REV. 1425, 1436–37 (1968) (“This way of putting the matter acknowledges that the required mode of culpability may not only vary from crime to crime but also from one to another element of the same offense—meaning by material element an attribute of conduct that gives it its offensive quality”); see also Ronald L. Gainer, \textit{The Culpability Provisions of the Model Penal Code}, 19 RUTGERS L.J. 575, 577 (1988) (describing element analysis as the Model Penal Code’s greatest achievement). But for an early recognition of this framing, see Francis Bowes Sayre, \textit{The Present Signification of Mens Rea in the Criminal Law}, in \textit{Harvard Legal Essays} 399, 404 (Roscoe Pound ed., 1934) (“An intelligent understanding of the various states of mind requisite for criminality can be gained only through an intensive study of the substantive law covering each separate group. The old conception of mens rea must be discarded, and in its place substituted the new conception of mentes reae.”).

\textsuperscript{50} Or what is also referred to as the “elemental,” “descriptive,” and “narrower” accounts of mens rea. See Garvey, supra note 6, at 546.

Failing to clearly recognize these distinctions has historically been the cause of much mischief in the criminal law, which the narrow approach has gone a long way toward eliminating.\(^{52}\) Indeed, the PKRN view truly revolutionized—in the best way possible—statutory and judicial treatment of mens rea in American criminal law.\(^{53}\) However, the benefits that accrue from thinking about mens rea in these narrow terms in the legal context should not lead one to view these modern distinctions as more than they are—namely, a helpful but artificial construct. And while the PKRN understanding may be practically useful for those tasked with clearly drafting or interpreting criminal laws, viewing mens rea as only PKRN may lead to problems of its own. For example, taking the PKRN view may obscure the moral justification for conditioning blame upon mental states in the first place. And it may also lead us to miss important ways that the criminal law is failing to respect the guilty mind in practice. In which case limiting ourselves to this narrow understanding could ultimately yield a view of mens rea reform—both our need for it as well as its potential manifestations—that is less persuasive or relevant than it otherwise could be.

Allow me to illustrate with an example of the kind of policy issue that the narrow view of mens rea might lead one to overlook. Consider the possibility of intentional wrongdoing which, although unjustifiable or excusable in conventional terms, is accompanied by a state of mind that is insufficiently blameworthy to justify a criminal conviction.\(^{54}\) The two felony convictions for assault of a police officer (APO) at issue in *Crossland v. United States*—a 2011 case arising out of the District of Columbia—are illustrative.\(^{55}\)

At the time the assaults in question occurred, the defendant—a young black man—was mowing his lawn on Saturday and talking to his cousin.\(^{56}\) Although the police were well aware that the defendant and his friend were not doing anything unlawful, they nevertheless chose to subject the two of them to a series of “aggressive” and indisputably “unconstitutional” patrols that were—quite regretfully—department practice at the time.\(^{57}\) Lacking “any right to go up and start searching them,” the officers aggressively “went up and seized them, told them to turn around, and started patting them down.”\(^{58}\) The defendant “initially complied, but quickly became agitated,

\(^{52}\) For an unfortunate recent example of failing to recognize the difference between these two understandings, see Justice Kagan’s opinion for the Court in Kahler v. Kansas, 140 S.Ct. 1021, 1032 (2020) (“Early commentators on the common law proposed various formulations of the insanity defense, with some favoring a morality inquiry and others a mens rea approach.”); compare id. at 1042 (Breyer, J. dissenting) (“The Court points out, correctly, that many of the common-law sources state that the insane lack mens rea or felonious intent. But what did they mean by that? At common law, the term mens rea ordinarily incorporated the notion of “general moral blameworthiness” required for criminal punishment . . . . The modern meaning of mens rea is narrower and more technical.”); Joshua Dressler, *Kahler v. Kansas: Ask the Wrong Question, You Get the Wrong Answer*, 18 OHIO STATE JOURNAL OF CRIMINAL LAW (forthcoming 2020) (“Kagan was saying that Kansas’s statutory system, which permits a defendant to introduce mental health evidence to disprove the mens rea element of a crime (here, Kahler’s intent to kill), is consistent with one common law version of the insanity defense. This historical analysis is wrong or, at least, overstated . . . .”).

\(^{53}\) See, e.g., Francis X. Shen et. al., *Sorting Guilty Minds*, 86 N.Y.U. L. REV. 1306, 1315-16 (2011) (PKRN framework “accomplished what no legal system had ever expressly tried to do: orchestrate the noise of culpability into a reasonably uniform and workable system.”)

\(^{54}\) For general discussion of and the differences between justifications and excuses, see supra note 44.


\(^{56}\) Id. at 1009.

\(^{57}\) Id.

\(^{58}\) Id.
telling [the police] words to the effect of ‘Fuck this shit. I’m tired of this.’”

What happened next is disputed, but it appears that the defendant intentionally elbowed one of the officers in the head, at which point all of the officers present beat the defendant with fists and pepper spray. During the melee, the defendant briefly resisted the attempts of one of the officers to place him in handcuffs. On these facts, the defendant was convicted in a bench trial of two counts of APO—one premised on the elbow, the other on the momentary resistance.

The court’s determination in this case was based on its acceptance of the officers’ questionable accounting of the events in question. But even accepting the government’s version of how things transpired, one can reasonably ask: did the defendant possess mens rea? From the narrow, PKRN perspective the answer is clearly “yes.” The defendant purposely struck the first officer, and he purposely resisted the efforts of the second officer to arrest him. That’s enough to satisfy the government’s affirmative burden of proof with respect to the mental state requirement governing the offense in question.

Nor, does it seem, that an affirmative justification or excuse defense was available to the defendant. For one thing, the statute in question categorically stated that “[i]t is neither justifiable nor excusable cause for a person to use force to resist an arrest when such an arrest is made by an individual he or she has reason to believe is a law enforcement officer, whether or not such arrest is lawful.” But even absent this express legislative declaration, it does not seem that either of these affirmative defenses would have succeeded under the circumstances. For example, it seems clear that the defendant’s resistance to an unlawful search risked greater harm from the resulting confrontation than could have reasonably been prevented by the force he employed, in which case his conduct was unjustifiable. Nor would the defendant’s conduct have been excused under

---

59 Id. at 1007.
60 Id.
61 Id. at 1008.
62 As the DCCA observes:

The trial court specifically credited Officer Baldwin’s testimony, which was to the effect that appellant “delivered the elbow strike” and attempted to hit the officer while in a “fighting stance” (i.e., that appellant’s assaultive conduct toward the officer was intentional); that appellant did so without the officer having used any force against appellant; and that Officer McCue jumped on appellant’s back and struck his knees only after appellant had attempted to hit Officer Baldwin. The court also credited Officer Baldwin’s testimony that appellant resisted Officer Baldwin's and Officer McCue's efforts to handcuff him.

63 In contrast, Crossland testified “that he did not elbow or try to hit Officer Baldwin,” but that the “officer tried to throw him to the ground and then punched him in the eye when, instead of immediately sitting down as the officer instructed him to do after the pat-down, [Crossland] asked whether he could pull up his pants.” Id. at 1007.
64 Again, this paper is working with the version of the facts that the court accepted, which may not be an accurate representation of what actually happened.
65 See, e.g., Craig Mackey, Hudson v. Michigan and the Ongoing Struggle for Accountability in Law Enforcement Institutions, 6 ALB. GOV’T L. REV. 606, 631 (2013) (“If citizens who believe they are being wrongfully arrested will calmly submit to police officers who are detaining them, there will be fewer violent confrontations, and less people will get injured and killed.”); In re C.L.D., 739 A.2d 353, 355 (D.C. 1999) (noting that the policy behind D.C.’s APO statute is “to deescalate the potential for violence which exists whenever a police officer encounters an individual in the line of duty”).
traditional principles of duress, given that the officers’ unlawful search did not threaten the amount of violence—death or severe bodily injury—typically required for this affirmative defense.66

And yet, I think that approaching the question of “did the defendant possess mens rea” from the broader guilty minds perspective suggests the possibility of a different conclusion. To understand why, try putting yourself in the defendant’s situation: you are engaged in completely peaceful activities when you’re swarmed by an aggressive group of officers who work for a department that has visibly and systematically violated the constitutional rights of people in your community. Most likely, you know people who have already been subjected to this kind of harassment (or worse) before, and you can certainly expect that you or others will face similar forms of harassment again in the future. Overcome with a sense of anger, frustration, and hopelessness, you lose control and lash out, at first violently and then later by briefly attempting to frustrate the arresting officer after having been subjected to a group beating. Viewed in this light, it’s easy understand why the defendant responded the way he did—indeed, his response might be perceived as reasonable under the circumstances.

Which is not say what the defendant did was right or good, in which case his conduct might constitute a justified act of self-defense. Nor is it to say that the defendant was wholly coerced, or “so intimidated that he was unable to choose otherwise,” in which case he might plausibly have a duress defense in at least some jurisdictions.67 There’s little doubt, I think, that the defendant’s response was less a matter of intimidation than it was one of frustration and anger. But even that could make it difficult to conclude that the defendant was sufficiently blameworthy to justify two serious felony assault convictions.68 This is so on the facts presented, which raise the possibility that the defendant’s conduct was solely attributable to his morally appropriate anger and frustration with the police’s unquestionably illegal and abusive behavior.69 Yet the guiltiness of the defendant’s mind might be all the more dubious in the face of what the court might have learned if it had taken the time to inquire.

66 For example, as Robinson et al. explains, supra note 43, the majority approach to duress requires the defendant to have been coerced by what he “reasonably believed was an unlawful threat of imminent death or severe bodily injury to himself or another.”
67 MODEL PENAL CODE § 2.09, cmt. at 373 (MPC approach to duress focuses on a situation where “actor makes a choice, but claims in his defense that he was so intimidated that he was unable to choose otherwise”); see MODEL PENAL CODE § 2.09(1) (framing the test for duress in terms of whether the defendant “was coerced to [engage in the offense] by the use of, or a threat to use, unlawful force against his person or the person of another, that a person of reasonable firmness in his situation would have been unable to resist”); Id., cmt. at 376 (“The typical situation in which the section will be invoked is one in which the actor is told that unless he performs a particular criminal act a threatened harm will occur and he yields to the pressure of the threat, performing the forbidden act.”); see also Robinson et al., supra note 43 for discussion of relevant legal trends.
68 At the time, the District’s APO statute provided a 5-year statutory maximum for commission of the offense. D.C. Code § 22–405(a) (2001) (“Whoever without justifiable and excusable cause, assaults, resists, opposes, impedes, intimidates, or interferes with any officer or member of any police force operating in the District of Columbia, ... while engaged in or on account of the performance of his or her official duties, shall be fined not more than $5,000 or imprisoned not more than 5 years, or both. It is neither justifiable nor excusable cause for a person to use force to resist an arrest when such arrest is made by an individual he or she has reason to believe is a law enforcement officer, whether or not such arrest is lawful.”)
69 Crossland v. United States, 32 A.3d 1005, 1008 (D.C. 2011) (“We discern no reason to doubt (and the government does not dispute) that Officer Baldwin’s conduct—forcibly searching appellant when, as the officer acknowledged, appellant was doing nothing unlawful—violated appellant’s Fourth Amendment right to be free from unreasonable searches and seizures.”).
For example, maybe the defendant (or a close friend or relative) had experienced similar police abuses in the past—and observed, with tremendous frustration, the government officials getting away with it time and time again. Perhaps the trauma of these unjust and unlawful experiences spilled over into other parts of his life, harming his relationships, employment prospects, and ultimately, his ability to live a meaningful life. Or consider the possibility that the defendant’s decision to lash out was also influenced by a cognitive issue he was suffering from at the time—for example, deep depression or some form of psychological impairment.

In each of these situations, the defendant may have found it all the more difficult to do what the law asked of him under the circumstances. True, none of these impairments would likely rise, individually, to the level of a general excuse defense, which would categorically exclude him from criminal liability. But a more holistic analysis of the defendant’s state of mind may still make it difficult to conclude that his conduct was a product of his failing to care about the law, or the interests that an APO statute protects (i.e., the physical security of police), as opposed to situational factors well beyond the defendant’s control.\footnote{See infra Section B for further discussion of these issues.} In that case, one is left with serious reservations about the foundational question presented by mens rea in the broadest sense: was the defendant’s state of mind sufficiently blameworthy to support a finding of guilt under the circumstances?

This is just one example of how the guilty minds understanding of mens rea might do a better job of opening our eyes to areas of the criminal law that merit careful examination in a way that the narrower PKRN approach might otherwise miss. So it’s important to keep this broader approach in mind when thinking about or discussing possibilities for mens rea reform. However, before committing oneself to a program focused on the guilty mind, it’s worth taking some time to consider two very basics question: (1) why do we care about the guilty mind; and (2) why should we care about the guilty mind? Section B, below, addresses the first question, by exploring the ways in which we account for the guilty mind in our everyday moral assessments of blame and responsibility. Thereafter, Section C addresses the second question, making the normative case for conditioning our private and criminal law judgments of blame upon the guilty mind.

B. Why Do We Care About The Guilty Mind?

Why do we care about the guilty mind? There are many ways to approach the question, but I think the best place to start is with the following observation: much of what we care about in life is rooted in the mind and the decisions that arise from it.\footnote{See, e.g., Stephen J. Morse, Criminal Law and Common Sense: An Essay on the Perils and Promise of Neuroscience, 99 MARQ. L. REV. 39, 52 (2015) (“Virtually everything for which agents deserve to be praised, blamed, rewarded, or punished is the product of mental causation and, in principle, is responsive to reasons, including incentives.”).} This is particularly true when it comes to evaluating ourselves and other conscious creatures.\footnote{See, e.g., id. (“Responsibility judgments depend on the mental states that produce and accompany our bodily movement and stillness. This is how we think about ourselves, and this is the concept of the person that morality and law both reflect.”).} For example, we celebrate good decisions, condemn bad decisions, and withhold judgment for those—such as infants, pets, and the severely disabled—whose ability to make decisions is limited or diminished.\footnote{See, e.g., sources cited supra notes 11 and accompanying sources.} This emphasis on the minds of others serves as a cornerstone of our emotional lives, the interpersonal
relationships we form, and the societies we build together. It also says something important about what it is to be a human: we are likely the only creatures who experience physical movements directed by a mind as having meaning or saying something important, both about the act and the actor.

Consider, for example, the difference between hearing that “John burned the Sin City movie poster he had hung in his dorm room because he thought doing so would impress his fraternity brothers” and hearing that “John burned the Sin City movie poster he had hung in his dorm room in order to protest the movie’s sexist displays of women and Harvey Weinstein’s involvement.” The act in both situations is the same; however, the objectives and priorities driving them are quite different. And that makes a world of difference both in terms of how we understand the bodily movement and perceive the person who made it. Simply put, take a look at the mind behind an act and you may very well encounter a story—about what the act means and what the actor cares about.

74 See, e.g., Peter Strawson, Freedom and Resentment, in PERSPECTIVES ON MORAL RESPONSIBILITY 45 (John Martin Fischer & Mark Ravizza eds., 1993); R. JAY WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS (1994); Strawson, supra note 12; see also Erik Luna, Spoiled Rotten Social Background, 2 ALA. C.R. & C.L.L. REV. 23, 33 (2011) (“reactive-attitude theory argues that moral responsibility is a function of the attitudes and emotions we feel in our interpersonal relationships, constituted in a community and creating expectations of all members”).

75 No doubt, as Justice Holmes famously observed, even a dog intuitively differentiates between a stumble and a kick. But it is only humans, by virtue of their innate sensitivity to the mental states of others and the circumstances that influence them, who understand the (counterintuitive, moral) difference between the disrespectful stumble of someone who couldn’t care enough to look out for another person and the beneficent kick of someone trying to move that same person from being struck by oncoming traffic. See, e.g., Stephen J. Morse, Inevitable Mens Rea, 27 HARV. J.L. & PUB. POL’Y 51, 62 (2003) (“[O]nly people create meaning and care about meaning and that these are further mental phenomena that are motivated and motivating.”); Morse, supra note 71, at 52. (“Machines do not deserve praise, blame, reward, punishment, concern, or respect because they exist or as a consequence of the results they cause. Only people, intentional agents with the potential to act, can do wrong and violate expectations of what they owe each other.”).

76 Consider Elizabeth Anderson and Richard Pildes:

The expression of a mental state brings that state into the open, for oneself and potentially for others to recognize. People recognize the mental state in its expression. This is an interpretive activity. To interpret what a statement means, we try to grasp what the speaker is saying. To interpret what an action means, we try to identify what the agent is doing. Deeds are identified, not by mere physical descriptions of bodily movement, but by the intentions that they express and that give them meaning. Interpretation is a matter of making sense of the speech or action in its context. Suppose an individual burns a piece of paper. What does this mean? If the paper is a draft card, and he burns it in the context of others doing the same thing at an antiwar rally, we understand his action to express outrage at the draft. See also SCANLON, supra note 10 (observing that “burning a Picasso painting because of a dare carries a very different meaning from my burning a Picasso painting because of how cruelly it represents women”).

77 Or for a simpler example: a push, a brush, a grope, a swipe, and a caress may all involve similar movements. But the intentions behind them are quite different, and for that reason we perceive them and the actor differently.

78 See, e.g., Kimberly Kessler Ferzan, Holistic Culpability, 28 CARDOZO L. REV. 2523, 2533 (2007) (“Mental states are essential ingredients in meaning.”); Morse, supra note 75, at 62 (“Mental states are thus vital because full understanding of action itself and an action’s moral significance depends upon the meaning of the action. Mental states signal both that what the agent has done is wrong and how wrong it is.”).
So it should come as no surprise that when one person harms another person—or otherwise engages in seriously wrongful conduct—we take a keen interest in what was happening in the mind of the actor.\[^{80}\] We instinctively gravitate toward things like intentions, motivations, awareness, and beliefs, all of which provide the most direct perspective on what this individual may have been thinking at the moment in question. But we also concern ourselves with other mental phenomena—for example, rationality and volitional control—from which we’re able to determine whether the wrongdoing was perpetrated by someone with the ability to think and act morally, in contrast to, say, a small child or someone suffering from serious mental illness.

An illustration will be helpful. Imagine you have undertaken the yeoman’s task of managing the entrance to a Costco that has just received its only shipment of toilet paper for the entire summer due to the supply chain disruptions resulting from a global pandemic. You’re in Phoenix, so although it’s only 10 am, it’s already 117 degrees and sweltering. Right as you’re about to let a young mother holding a young child through the sliding doors into the air conditioning, a large man comes from out of nowhere and jostles them, knocking both the mother and child over and onto the ground. Almost reflexively, you exclaim: “Hey, buddy, what the $@&% are you thinking?”

In posing this question, you are unlikely to be interested in all of the things that might have been on his mind—for example, his love of the band Creed,\[^{81}\] the smell of Lush soap stores,\[^{82}\] or the Transformers movie franchise.\[^{83}\] Rather, what you’d likely want to know more about are those aspects of his state of mind that are of moral relevance to the offensive act in question. For example, did he intend to knock the mother over? Was he aware that the mother was holding a child? What was he hoping to achieve by jumping the line? And was his decision the product of a mind capable of rational reflection and free choice—or was it distorted by the influence of psychotic thinking or a heat-induced emotional breakdown?

The forms of mental phenomena that grab our attention are varied and complex—but their moral salience is often obvious.\[^{84}\] And I will also suggest that they possess an intelligible structure,

\[^{80}\] See, e.g., Ferzan, supra note 79, at 2532 (“[T]here is a constitutive relationship between the internal states of the actor and our assessment of moral blameworthiness.”).
\[^{83}\] Cf. Yohana Desta, Transformers: The Last Knight Is a Million God-Awful Movies Crammed Into One, VANITY FAIR (June 20, 2017), https://www.vanityfair.com/hollywood/2017/06/transformers-the-last-knight-review
\[^{84}\] As Shoemaker observes:

We have many emotional responsibility-responses to ourselves and others, including admiration, disdain, shame, pride, regret, disappointment, approval, anger, and gratitude (among many others). Much of the time, we may feel that this entire range of responses is available for the agents we come across, but for some agents we hesitate, feeling that only some of these responses are appropriate, whereas other responses would be inappropriate. This is true, I suggest, of some people with mild intellectual disabilities, including also those at the high-functioning end of the autism spectrum, people with mild-to-moderate Alzheimer’s dementia, those with depression, those with obsessive-compulsive disorder, psychopaths, and those from morally-deprived upbringings.

Shoemaker, supra note 10, at 911–12.
which is comprised of four components, or areas of inquiry, the importance of which our minds intuitively grasp. In what follows, I offer one way to describe this structure.

Question No. 1: Why did the person do it? The first, and perhaps most powerful, domain of psychological interest is focused on understanding an actor’s reasons for engaging in wrongful conduct. Consider, for example, how you might feel upon learning that the line-cutter’s jostle was motivated by one of the following:

(1) A desire to send a message of racial inferiority to the young woman (and her child); or

(2) A desire to ensure that the line-cutter and his family would have toilet paper for the summer; or

(3) A desire to help the woman and her child avoid what he mistakenly perceived to be a car about to sideswipe them.

Question No. 2: Was the person aware of the risks involved? The second domain of psychological interest is focused on understanding the nature of an actor’s awareness of the risks associated with his or her wrongful conduct. Our interest tracks two primary dimensions: (1) the

---

85 See also Serota, supra note 46, at 1204-05 (alluding to this structure).
86 For criminal law theory emphasizing the importance of motivations to moral responsibility, see Alexander, Ferzan, Binder, and many of the other sources cited supra note 4; see also Scanlon, supra note 10 (“The agent’s reasons for acting (and the fact that other considerations did not count for him as reasons against so acting) are what constitute his attitude toward others, and what have the implications that blame involves[.]”).

For empirical work suggesting that the public holds similar views (and that motivations may be most significant to perceived blameworthiness), see PAUL H. ROBINSON & JOHN M. DARLEY, JUSTICE, LIABILITY, AND BLAME (1995) (Studies 5, 6, 8, 9, and 16); Janice Nadler & Mary-Hunter McDonnell, Moral Character, Motive, and the Psychology of Blame, 97 CORNELL L. REV. 255 (2012); Paul H. Robinson & John M. Darley, Testing Competing Theories of Justification, 76 N.C. L. REV. 1095 (1998); see also PAUL H. ROBINSON, INTUITIONS OF JUSTICE AND THE UTILITY OF DESERT 313-14 (2013) (finding across a number of studies involving accomplice liability that people assign more liability and punishment for purpose, as opposed to knowledge or recklessness, as to facilitating criminal conduct).

It’s important to acknowledge that the criminal law often claims that “motivation” is immaterial to liability and punishment. See, e.g., Elaine M. Chiu, The Challenge of Motive in the Criminal Law, 8 BUFF. CRIM. L. REV. 653, 656 (2005) (“Generations of scholars of the criminal law have learned that motive is irrelevant in the criminal law.”); Douglas Husak, Motive and Criminal Liability, 8 CRIM. JUST. ETHICS 3, 3 (1989) (“This thesis is endorsed, sometimes with minor qualifications, by almost all leading criminal theorists.”). But this at best misleading and more likely just plain false: from justification defenses, to bias crimes, to the gross deviation standards governing the common definitions of recklessness and negligence, to sentencing decisions, the criminal law accounts for motivations in multifarious ways. See, e.g., Carissa Byrne Hessick, Motive’s Role in Criminal Punishment, 80 S. CAL. L. REV. 89 (2006). And that’s to say nothing of the informal ways that motive likely influences the decisions of courts, prosecutors, and juries. See, e.g., Anna Roberts, Dismissals as Justice, 69 ALA. L. REV. 327, 376 (2017) (providing illustrations in the context of de minimis dismissals). This should not be surprising, however, given that a prohibition on considering motive “violates widely held intuitions.” Binder, supra note 4, at 426; see also Kimberly Kessler Ferzan, Plotting Premeditation’s Demise, LAW & CONTEMP. PROBS., (2012), at 83, 93–9 (“If Alex decides to drive one hundred miles per hour on the highway, whether society deems Alex culpable and deserving of blame and punishment will depend upon whether he has chosen to impose this risk to impress his friends with how fast his car can drive or, alternatively, to transport a critically injured friend to the hospital.”).

87 For philosophical literature emphasizing the importance of awareness of risk to moral responsibility, see, for example, Alexander, Ferzan, Simons, Yaffe, Binder, cited supra note 4; see also Ferzan, supra note 86 at 2532 (“That is, we judge David to be more or less culpable depending upon whether he knew Vic was there, thought there was a
level of awareness possessed by the wrongdoer (e.g., awareness that harm was possible versus awareness that harm was nearly certain to occur); and (2) the gravity of the harm of which the actor was aware (e.g., awareness that a serious injury was possible versus awareness that a minor injury was possible). Consider, for example, how you might feel upon learning that the line-cutter:

(1) was aware that his conduct would almost surely cause serious injury to both the mother and child; or

(2) was only aware that his conduct might (at worst) cause a minimal injury to both the mother and child; or

(3) was completely unaware that his conduct posed any risk of injury to either the mother or child.

Question No. 3: Was the person able to engage in moral reasoning? The third domain of psychological interest is focused on understanding the extent to which an actor possessed the ability to engage in moral reasoning (i.e., the ability to distinguish right from wrong) when he or she engaged in the wrongful conduct. Consider, for example, how you might feel upon learning that the line-cutter:

chance Vic might be there, or believed Vic was not there. Without knowing the risks David understood his action to be imposing, we cannot attribute meaning to that action.”).

For empirical work suggesting that the public holds similar views, see ROBINSON & DARLEY, supra note 86 (Studies 5, 6, 8, 9, and 16); ROBINSON, supra note 86, at 305–06, 313–14, 327.

Note, however, that two recent studies call into question (1) whether lay jurors can reliably make fine-grained distinctions as to the extent of an actor’s conscious risk awareness and (2) whether material variances in conscious risk awareness, even when accurately identified by lay jurors, are viewed as increasing an actor’s blameworthiness. Matthew R. Ginther et al., The Language of Mens Rea, 67 VAND. L. REV. 1327, 1339–61 (2014) (finding “conflation of K and R punishment because subjects do not see a clear moral distinction between the K and R mental states, at least as it concerns the result element of offenses.”); Francis X. Shen et al., Sorting Guilty Minds, 86 N.Y.U. L. REV. 1306, 1326–54 (2011); see also Francis X. Shen, Minority Mens Rea: Racial Bias and Criminal Mental States, 68 HASTINGS L.J. 1007, 1017–42 (2017) (conducting a study reaffirming that lay jurors struggle to distinguish between material variances in an actor’s conscious risk awareness but also finding that culpable mental state evaluations of this nature may be resistant to implicit racial bias).

For a fascinating study employing neuroimaging and machine-learning techniques to explore the neural correlates of variances in conscious risk awareness, see Iris Vilares et al., Predicting the Knowledge-Recklessness Distinction in the Human Brain, 114 PROCEEDINGS NAT’L ACADEMY SCI. 3222 (2017).

There are many other dimensions, such as, for example, the difference between believing something to be true, which does not in fact turn out to be true. I leave out these complexities in the interests of accessibility.

For philosophical literature emphasizing the relationship between blameworthiness and the extent to which an agent possesses the rational capacities necessary to distinguish right from wrong, see, for example, ANTONY DUFF, ANSWERING FOR CRIME: RESPONSIBILITY AND LIABILITY IN THE CRIMINAL LAW 39–40 (2007); David O. Brink & Dana K. Nelkin, Fairness and the Architecture of Responsibility, in OXFORD STUDIES IN AGENCY AND RESPONSIBILITY 285 (David Shoemaker ed., 2013); Stephen J. Morse, Excusing and the New Excuse Defenses: A Legal and Conceptual Review, 23 CRIME & JUST. 329, 340 (1998); Alexander & Ferzan, supra note 4 at 155 (“Rationality is the cornerstone of responsible agency. If an actor cannot comprehend or respond to norms, then it cannot be said that laws or morality are properly addressed to the actor.”).

For empirical work suggesting that the public holds similar views, see ROBINSON & DARLEY, supra note 86 (Studies 12 & 13); ROBINSON, supra note 86, at 342 (“Perpetrators who are judged to be suffering from a high degree of dysfunction, whether that dysfunction is of the cognitive or conduct control sort, are normally not assigned criminal liability . . . Typically, more than 70 percent of the respondents assign no criminal liability to such cases.”); id. at 348
(1) was experiencing extreme psychotic delusions, which made it impossible for him to recognize that injuring the mother and child was immoral; or

(2) was experiencing low-level psychotic delusions, which made it meaningfully more difficult for him to recognize that injuring the mother and child was immoral; or

(3) did not experience any diminishment in his ability to engage in moral reasoning, or distinguish right from wrong.90

Question No. 4: Was the person able to control him or herself? The fourth, and final, domain of psychological interest is focused on understanding the extent to which an actor possessed the ability to control him or herself, morally speaking, when he or she engaged in the wrongful conduct.91 Consider, for example, how you might feel upon learning that the line-cutter, although believing what he did to be wrong, was:

("[B]oth cognitive and control dysfunction appear to support a defense" for involuntary intoxication); James R.P. Ogloff, A Comparison of Insanity Defense Standards on Jury Decision Making, 15 LAW & HUM. BEHAV. 509, 513 (1991); Norman J. Finkel, Insanity Defenses: From the Jurors' Perspective, 9 LAW & PSYCHOL. REV. 77, 79 (1985); see also Stephen J. Morse, Reason, Results, and Criminal Responsibility, 2004 U. ILL. L. REV. 363, 440 ("[T]he capacity for rationality is a congeries of skills, including the ability to perceive accurately, to reason instrumentally according to minimally coherent preference-ordering, and to appreciate the significance of reasons and their connection to our actions.").

Arguably, there exists a related, but distinct dimension not directly addressed by this, or the other three questions: whether and to what extent the person was aware that his or her conduct was morally or legally prohibited. See, e.g., Garvey, supra note 6, at 556–57 ("Our reactive emotions tend, all else being equal, to subside if we come to believe that a wrongdoer (including a criminal wrongdoer) didn’t realize he was doing wrong, which is at least some evidence that ignorance has the power to excuse. Or maybe it would be more accurate to say that the retreat of the reactive emotions in the face of ignorance is some evidence that ignorance does excuse in morality and should excuse in criminal law."); HUSAK, supra note 39; compare Gideon Yaffe, Excusing Mistakes of Law, 9 PHILOSOPHERS’ IMPRINT 1, 8 (2009) ("To say that [someone who intentionally kills with full awareness of all the facts] also needs to believe himself to be acting illegally if he’s to be doing something deserving of punishment is to elevate a reverence for the law beyond morally tolerable bounds."). As mentioned supra note 39, I omit this aspect largely for purposes of accessibility.

90 This example analyzes the relationship between blame and mental illness in terms of “mental illness’s effect on the actor’s ability to appreciate the wrongness of imposing the risks for the reasons he has for doing so.” Kimberly Kessler Ferzan, Patty Hearst Reconsidered: Personal Identity in the Criminal Law, 15 OHIO ST. J. CRIM. L. 367, 382 (2018). However, mental illness can also have an “effect on the actor’s ability to refrain from acting culpably – that is, by its effect on his volitional control.” Id. Question No. 4, below, explores this relationship but as it relates to involuntary intoxication, instead of mental illness.

91 For philosophical literature emphasizing the relationship between blameworthiness and the extent to which an agent possesses the volitional capacities necessary to conform one’s conduct to his or her moral judgment, see, for example, Brink & Nelkin, supra note 89; HART, supra note 4, at 23; Paul H. Robinson, A System of Excuses: How Criminal Law’s Excuse Defenses Do, and Don’t, Work Together to Exculpate Blameless (and Only Blameless) Offenders, 42 TEX. TECH L. REV. 259, 264 (2009); Dressler, supra note 42, at 258; Matthew Talbert, Implanted Desires, Self-Formation and Blame, 3 J. ETHICS & SOC. PHIL. 2, 5 (2009) ("[T]he question of whether it is reasonable to blame Beth will be best answered by inquiring into whether she is capable of governing her behavior according to internal values and judgments so that her behavior expresses interpersonally significant attitudes.").

For empirical work suggesting that the public holds similar views, see ROBINSON & DARLEY, supra note 86 (Studies 12 & 13); ROBINSON, supra note 86, at 342 ("Perpetrators who are judged to be suffering from a high degree of dysfunction, whether that dysfunction is of the cognitive or conduct control sort, are normally not assigned criminal liability . . . Typically, more than 70 percent of the respondents assign no criminal liability to such cases."); Id. at 348 ("both cognitive and control dysfunction appear to support a defense" for involuntary intoxication).
(1) completely unable to control his conduct, which was an unexpected side effect of a prescription medication that his negligent doctor had failed to warn him about;

(2) substantially unable to control his conduct, which was an unexpected side effect of a prescription medication that his negligent doctor had failed to warn him about;

(3) fully capable of controlling his conduct, because he did not take the prescription medication.

That the diverse mental phenomena that comprise these four domains of inquiry seem to be of similar relevance to us is itself a fascinating fact about our moral lives. But even more impressive is the fact that we—again, without much effort—are able to synthesize them into judgments about blame and moral responsibility. To illustrate, consider three possible psychological profiles of the line-cutter, each of which arises from different combinations of the domain-specific mental states described above.

(1) The line-cutter was motivated by a desire to send a message of racial inferiority to both the mother and the community. At the time of the act, he believed that he would almost surely cause serious bodily injury to both the mother and child. He was not suffering from any cognitive or volitional impairments; rather, he was fully capable of engaging in rational reflection and was also in full control of his conduct.

(2) The line-cutter was motivated by a desire to secure for his family a supply of toilet paper. At the time of the act, he believed that his conduct posed a small risk of minimal injury to both the mother and child. He was also suffering from a moderate mental illness, which made it meaningfully more difficult for him to distinguish between right from wrong.

(3) The line-cutter was motivated by a desire to save the woman and child from what he mistakenly perceived to be a car about to run them over. At the time of the act, he perceived no risk of harm to either of them from his jostling. And even if he did, it wouldn’t have mattered because earlier in the day he took a prescription medication (for the first time), at the behest of his negligent doctor, which made it impossible for him to control his conduct.

After reading through these examples, many may find it easy—or at least easier than one might have otherwise thought—to answer a couple of questions. The first is binary (yes/no): are each of these actors blameworthy? And the second is comparative: how does the blameworthiness of each actor compare with the others on a single continuum of culpability? (Which is to say, among the three scenarios, which actor is: (i) most blameworthy; (ii) least blameworthy; and (iii) falls in between?)

---

92 For example, readers may have deemed the first actor to be: (1) clearly blameworthy and (2) more deserving of blame than the other two actors. This is in contrast to the actor in the third scenario, who readers may have deemed to be: (1) clearly blameless and (2) less deserving of blame than the other three actors. Evaluation of the second actor may have been a bit more difficult—for example, some may have perceived the case for blameworthiness to be
finding, for example, that the first actor clearly is blameworthy, that the third actor clearly is not, and that each of the three psychological profiles rests on a continuum of descending blameworthiness (i.e., the first actor is more blameworthy than the second who is more blameworthy than the third).

Simply put, this exercise attempts to illustrate what a number of psychological studies have found: (1) we have a relatively refined sense of whether the mind behind an act is blameworthy; and (2) we have an even clearer sense of comparative psychological blameworthiness (i.e., how differing states of mind compare to one another on an ordinal blameworthiness scale).93

What are we to make of all this? There are many interesting things that might be said about our shared capacity for rendering moral responsibility judgments premised on an assessment of someone’s state of mind.94 However, for purposes of this paper, I’d like to focus on a narrow, though still complex (and speculative), dimension of these psychological findings: that these mental state evaluations evidence a certain level of *moral coherence*—which is to say, that they’re logical and consistent when viewed in light of our moral commitments.

That this would be the case is far from obvious given the evolutionary forces that seem to have shaped our moral intuitions.95 So it would be unsurprising if the way we blame other minds turned out to be as seemingly haphazard or incongruous as those evolutionary forces often appear. But upon closer scrutiny, what one finds is a unifying thread that does a pretty good job of making moral sense of them. That thread is reflected in a principle that a number of contemporary moral and criminal law theorists have highlighted in their work—what I will refer to as the principle of insufficient concern.96

---


95 See, e.g., Robinson & Kurzban, supra note 94; Hoffman, supra note 94; see also Stephen J. Morse, *Thoroughly Modern: Sir James Fitzjames Stephen on Criminal Responsibility*, 5 OHIO ST. J. CRIM. L. 505, 513–14 (2008) (“Social constructivists might object to this account of essential human nature, but recent empirical evidence supports the claim that we are predisposed to be retributivists and that the social emotions generally are part of our innate repertoires.”).


For important contributions from the realm moral philosophy see, for example, JOHN FISCHER AND MARK RAVIZZA, *Responsibility and Control: A Theory of Moral Responsibility* (1998); Angela Smith,
The principle of insufficient concern is typically understood to constitute a normative principle about which minds we ought to deem blameworthy; however, I believe it also does a reasonably good job of making sense of how we do, in fact, blame other minds. At the heart of the principle is the idea that the mind is of moral significance because of what it reveals about a person’s values. By “values,” however, I don’t mean any values—or character traits—but rather, incident-specific values, which capture the evaluative weight someone afforded to his or her own interests in contrast to those of another person (or society).

Think of it this way: typically, someone who consciously disregards a risk of harm to another person (or to society) has effectively made a choice that the benefits flowing from his or her conduct outweigh its attendant costs. Sometimes, this choice is one that society can endorse; for example, even the most cautious surgeons, construction workers, and emergency responders will on occasion knowingly impose risks upon others. Yet we typically praise—rather than blame—them for doing so, provided they act reasonably and their reason for doing so is the greater good of people or society.

Responsibility for Attitudes: Activity and Passivity in Mental Life, 115 ETHICS 236 (2005); SCANLON, supra note 10; Pamela Hieronymi, Controlling Attitudes, PACIFIC PHILOSOPHICAL QUARTERLY 45 (2006).

It’s important to note that within this body of literature, there are many disagreements—for example, criminal law theorists dispute whether and to what extent negligent inadvertence can manifest the kind of insufficient concern necessary to ground criminal liability and punishment. Compare Kenneth W. Simons, Culpability and Retributive Theory: The Problem of Criminal Negligence, 5 J. CONTEMP. LEGAL ISSUES 365 (1994), with Alexander & Ferzan, supra note 4; see also Douglas Husak, Negligence, Belief, Blame and Criminal Liability: The Special Case of Forgetting, 5 CRIM L. & PHIL. 199 (2011).

In one sense, this is not at all surprising given the central role that moral intuitions play in the work of normative theorists. See, e.g., Douglas Husak, What Moral Philosophers Might Learn from Criminal Theorists, 36 RUTGERS L.J. 191, 192 (2004) (“Moral philosophers and criminal theorists tend to employ a similar methodology in deciding whether a given principle or doctrine is fair, just, or reasonable. Both attach extraordinary significance to our moral intuitions. They frequently begin by describing examples, and ask how the behavior of the persons in these cases should be assessed.”).

For discussion of the different ways in which this principle might be construed, see Kenneth W. Simons, Retributivism Refined—or Run Amok?, 77 U. CHI. L. REV. 551, 566 (2010) (“The principle of Insufficient concern can be understood in one of two very different ways. It could be a mere label, a term of art for unjustifiable and inexcusable acts. On this view, the actual standards for determining the actor’s culpability require us to look at more specific criteria of justification and excuse, of whether the actor lacks sufficient reasons for the risks he believes himself to be imposing. But, alternatively, insufficient concern could actually be an operative evaluative criterion of when acts are indeed unjustifiable and inexcusable.”). The discussion in this part tends towards the latter construction.

See, e.g., Gideon Yaffe, Intoxication, Recklessness, and Negligence, 9 OHIO ST. J. CRIM. L. 545, 553 (2012); Westen, supra note 4.

See, e.g., YAFFE, supra note 4 (“character requires stable dispositions of certain sorts, while what needs to be manifested in one’s act for culpability, on the view I am proposing, are a certain set of dispositions, whether or not they are stable”); Scanlon, supra note 10 (“Lack of ambition may be a fault of character, but in the account I am proposing it is not in itself grounds for blame by others. We might blame a person for his lack of ambition because this led him to let his family down, but this would actually be blame for violating his family obligation.”).

Note, for example, that a person may violate human interests, such as, for example, where D punches V, whether or not punching is legally prohibited. Or a person may violate societal interests, such as, for example, where D sells X a form of contraband, where the transfer of contraband is legally prohibited. See also YAFFE, supra note 16, at 78 (“It is perfectly possible to legally prohibit acts in virtue of features that have nothing whatsoever to do with the relevance of those acts to other people.”). Arguably, all legal interests (i.e., those protected by legal norms) are societal interests, but not all societal interests are legal interests.

See, e.g., YAFFE, supra note 4; Alexander & Ferzan, supra note 4.
The situation is entirely different, however, where such a person consciously disregards a risk—or perhaps fails to perceive one\textsuperscript{102}—for wholly selfish reasons. Consider, for example, the following situations:

(1) A doctor knowingly undertakes an unnecessary and dangerous surgery to make history.

(2) A construction worker uses a dangerous explosive at a construction site for the fun of it, when he knows a safer (but less entertaining) alternative exists.

(3) An emergency responder speeds someone to the hospital across busy city streets with his eyes glued to the radio instead of the road, because he cares more about good music than the safety of other people.

We intuitively view these individuals as blameworthy because they’ve placed their negligible personal gain—fame, fun, or the enjoyment of good music—over the very serious risks their conduct imposed upon the safety and well-being of others.

The principle of insufficient concern deems this entirely appropriate, in that each of these actors has failed to afford the interests of others the amount of weight in their decisionmaking that morality otherwise demands of them. Which is to say, in each of the above situations, the actor’s conduct manifests “a tendency to put himself first, to a greater degree than is acceptable.”\textsuperscript{103} In

\textsuperscript{102} The discussion in this Part largely focuses on conscious risk-taking; however, a similar analysis applies to unconscious risk-taking (i.e., culpable inadvertence), in which context we may be predisposed to blame those who unreasonably fail to perceive a risk of harm of which they should have been aware. See, e.g., Westen, supra note 4; Samuel Pillsbury, Judging Evil 171-172 (2012) (“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. Did the failure stem from a culpable lack of concern for the victim, or should we attribute it to other factors for which the individual should not be blamed?”). The principle of insufficient concern remains operative here, too, given relationship between what we value and what we attend to. As Angela Smith has observed:

\[W\]e take there to be some sort of rational connection between what we notice and what we evaluate or judge to be important or significant. We might characterize the connection in the following way: if one judges some thing or person to be important or significant in some way, this should (rationally) have an influence on one’s tendency to notice factors which pertain to the existence, welfare, or flourishing of that thing or person. If this is so, then the fact that a person fails to take note of such factors in certain circumstances is at least some indication that she does not accept this evaluative judgment.

Angela Smith, Responsibility for Attitudes: Activity and Passivity in Mental Life, 115 Ethics 236 (2005); see, e.g., Yaffe, supra note 98, at 565. Here’s an example. Imagine you bring your elderly parents to a baseball game during “take your parents to the ballpark day.” Sitting behind the batter’s box, a foul ball is hit in your direction. You immediately focus on protecting your parents from getting injured by the baseball, but pay no attention to the other elderly parents around you. The reason? You care more about your parents than these other people. The point? You are more likely to pay attention to what you care about.

\textsuperscript{103} As Gideon Yaffe phrases it:

\[W\]e care about an agent’s mental states, and the deliberative processes they guide, when assessing his responsibility because, thanks to them, his actions manifest his culpability-relevant values. In particular, thanks to the agent’s mental states, his actions manifest the evaluative weight that he gives to his own interests
that case, one would be justified in attributing the conduct to objectionable (moral) values, for which blame and moral responsibility seem appropriate, in contrast to arbitrary or external factors—for example, difficult circumstances or bad luck—for which they do not.

I will say more about this distinction in a moment, but first it’s important to explain why a theory of moral responsibility would place such great emphasis on caring for the interests of other people. This focus is indelibly connected to a basic tenet of our social morality: most of us believe, or at least live our lives under the assumption that, we all 

owe 

one another some degree of concern. No doubt, the contours of this obligation vary across relationships—most would presumably accept, for example, that we owe our children a significantly greater level of concern than, say, a random person on the street. But even to that random person’s children, no less than to the random person himself, some basic level of concern is owed—for example, to refrain from creating substantial risks of harm to their safety or well-being in the absence of strong countervailing reasons. And the place where that concern registers is the human mind.

To illustrate, consider another scenario involving the line-cutter from above. Imagine that the line-cutter is aware that his conduct might cause serious injury to both the mother and child, but that he also really wants to secure a summer supply of toilet paper. (His desire for toilet paper is, we might say, what motivates him.). Under these circumstances, someone who actually cares about the well-being of other human beings would view that awareness as a decisive reason to compare to the interests of other people. When a tendency to put himself first, to a greater degree than is acceptable, is manifested in his conduct, he is criminally culpable for that conduct.

YAFFE, supra note 1, at 25–26.

One thing to note: the discussion of the insufficient concern principle here more closely aligns with quality of will theorists, who focus on whether a wrongful act manifests an objectionable attitude towards others, in contrast to reasons-responsiveness theorists (which includes Yaffe), who focus on whether a wrongful act manifests something problematic about how the agent transacts with reasons. YAFFE, supra note 4, at 77. It is questionable whether any substantive difference between these two views exists. See id. (“To fail to grant the right degree of reason-giving weight to that fact is to manifest bad quality of will; and one cannot manifest bad quality of will of the relevant sort unless one also fails to grant appropriate reason-giving weight to the relevant fact about the act, namely that it causes pain.”). Nevertheless, I opt for the quality of will framing because it is arguably more intuitive and less awkward than speaking about “modes of transaction with reasons.” See also MICHAEL MCKENNA, CONVERSATION AND RESPONSIBILITY 57-64 (2012) (discussing and clarifying the meaning of the phrase “quality of will”).

104 See, e.g., Scanlon, supra note 10 (“Morality not only tells us to treat others in certain ways—to refrain from harming them, for example—but also gives us certain reasons for doing so. It tells us, for example, that their interests matter, and that we should take these interests as providing reasons. People who are indifferent to the interests of other rational beings are open to moral criticism on this account, whether or not they behave in ways that cause others harm or fail to help them in ways that they should.”); Id. (“[M]orality requires that we hold certain attitudes toward ward one another simply in virtue of the fact that we stand in the relation of ‘fellow rational beings’”).

105 See, e.g. Scanlon, supra note 10 (“Different relationships involve different [moral] standards.”); Strawson, supra note 12 (different relationship statuses demand different forms of “good will”).

106 This is reflected in the intuitive idea that a parent is obligated to take affirmative steps to protect one’s child from threats, even if it involves some level of danger to the parent—whereas one would not necessarily be obligated to do the same for some unrelated person’s children.

107 See, e.g., Scanlon, supra note 10 (morality “requires us to take care not to behave in ways that will harm those to whom we stand in this relation, to help them when we can easily do so, not to lie to them or mislead them, and so on”); YAFFE, supra note 4.

108 See, e.g., Alexander & Ferzan, supra note 4, at 7 (observing that blameworthiness is rooted in the “defendant’s decision to violate society’s norms regarding the proper concern due to the interests of others”); YAFFE, supra note 4.
refrain from engaging in the proposed course of conduct, based on the recognition that a mother and child’s interests in avoiding the risk of devastating—and potentially life altering—injuries are orders of magnitude more weighty than one person’s (or family’s) acquisition of a case of Charmin. So when he chooses to proceed anyway, thereby imposing the high risk of serious harm upon both the mother and child, it appears as though his wrongdoing is attributable to his failure to care enough about their well-being.

I say appears, because in order for us to safely draw a line between the line-cutter’s conduct and his failure to care enough about the mother and daughter’s well-being, we also need to know something about his ability to think and act morally—for example, was it entirely absent or meaningfully diminished? What I’m talking about here are the kinds of mental phenomena that could impair a person’s ability to engage in moral reasoning (i.e., determine right from wrong) or to control his or her conduct (i.e., conform one’s behavior to one’s moral judgments)—things like serious mental illness, immaturity, and severe emotional disruption. These rational and volitional capacities are generally considered to be a precondition for deliberating about and acting upon one’s values; therefore, to the extent they are diminished (or absent), so too is our basis for attributing wrongdoing to any deficiency in them. (Which is to say, “[i]f an individual has a mental disability, disturbance, or is overwhelmed by emotion, he cannot reflect meaningfully on

---

109 See, e.g., Ferzan, supra note 4, at 2532 (“Our evaluation that someone is culpable is a determination that that person’s action revealed insufficient concern for the interests of others, and to make this assessment we need to know the mechanics and the quality of the agent's reasoning.”); see also Morse, supra note 75, at 62 (“The mental states that accompany intentional bodily movements prima facie indicate whether the agent really did violate an expectation by intentionally moving his or her body and they thus prima facie indicate the agent’s attitude towards the rights and welfare of others.”).

110 See, e.g., Westen, supra note 4, at 366 (“An insane person who wrongfully kills another because of an unreasonable mistake of fact is excused because, although the insane person commits the actus reus of killing an innocent person, he lacks the attitudes of maliciousness, contempt, indifference, disregard, and neglect toward the legitimate interests of others that state-imposed blame represents offenders as possessing.”).

111 See, e.g., Westen, supra note 4, at 364 (“Because children are incapable of appreciating those interests in the way adults do, their conduct is incapable of manifesting disparaging attitudes toward those interests—or, at least, incapable of manifesting the kind of malice, contempt, indifference, disregard and neglect that the state expresses when it punishes criminal offenses.”).

112 See, e.g., Westen, supra note 23, at 151 (“Whether strong emotion consists of the urgent impulse of “fight or flight” that accompanies anger or the hopeless gloom of depression, strong emotion renders an actor less blameworthy because, while it does not make it impossible, it makes it much more difficult for an actor than if he were cool-headed to deliberate about and act upon his settled values.”); see also Ferzan, supra note 86, at 83 (“I also believe that some emotions do prevent us from having full access to our reasons.”).

113 Consider, for example, the import of discovering that the line-cutter was suffering from a deep psychosis at the time of his act that precluded his ability to recognize the moral difference between the mother and child’s interest in their bodily integrity versus his own interest in securing a summer supply of toilet paper. Or what if one was to learn that the line-cutter was forced to imbibe a narcotic at gunpoint a few moments earlier, which made it impossible for him to conform his behavior to what he might otherwise recognize as the morally right thing to do. In both of these scenarios, the line-cutter’s capacities of moral decision-making would be inhibited in a way that precludes us from attributing his conduct to a lack of sufficient concern for the mother and child. In contrast to a deficit of caring, the line-cutter’s conduct seems attributable to cognitive and volitional hurdles for which he bears no responsibility under the circumstances. See, e.g., Scanlon, supra note 10 (“It follows from the way in which blame depends on an agent’s reasons that conditions under which an agent acted, such as extreme stress or fear, can affect blame insofar as they affect the degree to which the action reflects the agent’s actual attitudes.”); Ferzan, supra note 86, at 83 (“If an individual has a mental disability, disturbance, or is overwhelmed by emotion, he cannot reflect meaningfully on his choices... When someone is very upset or angry, she may lack the ability to deliberate over her action in the way she otherwise would.”); Westen, supra note 4, at 156.
his choices,”114 in which case these impairments may “affect the degree to which the action reflects the agent’s actual attitudes.”115)

One straightforward way to understand this picture of moral responsibility is as follows: when wrongdoing occurs, blameworthiness hinges upon its basic source—objectionable values or bad luck. In the former situation, the person’s wrongdoing can be attributed to a deficit of concern for the interests of others, which falls below what society otherwise has reason to expect. In the latter situation, by contrast, the actor’s wrongful conduct is attributable to arbitrary circumstances that were thrust upon the individual. And the way that we—as outside observers—are able to make this determination is by evaluating a person’s state of mind in light of the four categories of mental phenomena described above.

In the final analysis, then, the principle of insufficient concern offers a rational accounting of the relationship between mental states and moral responsibility, which our wide-ranging and specific intuitions about psychological blameworthiness largely map onto.116 It is for this reason—namely, the general alignment of moral intuition with logical principle—that I believe our tendency to condition blame upon the guilty mind can be deemed coherent. With that, however, comes two important caveats.

Caveat No. 1: These are only patterns of moral judgment, which we can see reflected in a number of psychological studies, the philosophical literature on moral psychology, the legal literature on mens rea, the law, and perhaps in our own lives as well. But also, these sources are far from monolithic, as are the factors that influence our blameworthiness assessments.117 Indeed, it’s quite easy to find instances that conflict with the framework of mental state evaluations presented above. There are, for example, a number of situations in which our blameworthiness assessments seem to largely rest on the gravity of the harm caused or threatened, without regard to the nature of the mind accompanying it.118 And in other situations, our moral responsibility

---

114 Ferzan, supra note 96, at 83.
115 Scanlon, supra note 10.
116 That’s because, first, descriptively speaking they do a reasonably good job of making sense of what we do think makes a mind guilty. For example, all else being equal, it appears that:

(1) The worse an actor’s motivations, the greater his or her awareness of risk, the more blameworthy we are likely to perceive him or her to be; but also

(2) The greater the impairments to a person’s ability to think and act morally, the less blameworthy we are likely to perceive him or her to be.

And second, this pattern of moral evaluation generally tracks what the principle of insufficient concern would otherwise call for in any given situation.

117 For a good empirical accounting of these diverse sources, see Norman J. Finkle, Commonsense Justice: Jurors’ Notions of the Law (1995); see also Norman J. Finkel, Commonsense Justice, Culpability, and Punishment, 28 Hofstra L. Rev. 669, 701 (2000) (observing that “jurors are likely to flavor, combine, and cook them in more subjective and psychological ways, throwing in past experiences, intuitions, sentiments, biases, heuristics, construals, and prototypes, as they wok and roll”).

118 See, e.g., Carly Giffin & Tania Lombrozo, Wrong or Merely Prohibited: Special Treatment of Strict Liability in Intuitive Moral Judgment, 40 Law & Hum. Behav. 707 (2016); Joseph Sanders, et al., Must Torts Be Wrongs? An Empirical Perspective, 49 Wake Forest L. Rev. (2014). But note: even acknowledging this fact, it appears to be “evident, when the head-to-head evidence [between intent and harm] is examined, that hegemony goes to intent, as it
judgments seem to be principally influenced by morally irrelevant aspects of the person being judged, which range from the reprehensible—for example, race, class, religion, or sexual orientation—to the arbitrary—for example, which team the individual roots for on a Sunday afternoon. And yet, the fact remains that we can recognize these as departures precisely because there exists a coherent framework of mental state evaluation that is consistent with our basic moral commitments as well as what we think, say, and do about blame much of the time.

Caveat No. 2: Even if the way we apportion blame upon other minds is coherent, we may still have sound reasons to revise or discard the practice. That’s because coherence of a social practice is one thing, whereas its desirability is quite another. And when a social practice consistently leads to bad outcomes, we should not just acquiesce to the path of least resistance; rather, it’s important “to try to rise above our innate predispositions if they lead to morally [detrimental consequences].” Now, given how deeply rooted some predispositions may be, one could question how successful an attempt at disabusing ourselves of them would be. But as I argue in the next Part, our tendency to condition blame upon an assessment of the guilty mind is not something that we should resist (provided, that is, we opt to blame others at all). Rather, as I will argue, it is something that we have strong reasons to embrace, particularly for our criminal justice system. In what follows, I attempt to provide an accessible snapshot of those reasons, beginning first with the case of private blame, and then extending outward to the public blame inherent in criminal liability and punishment.

remains the starting point, and often the final point, in commonsense justice’s culpability analysis.” Norman J. Finkel & Jennifer L. Groscup, When Mistakes Happen: Commonsense Rules of Culpability, 3 PSYCHOL. PUB. POL’Y & L. 65, 117 (1997) (“Though the course does not run smooth or straight from beginning to end but contextually wends its way through related variables, those variables generally feed back into intent, through interaction effects, modifying, mitigating, and sometimes enhancing culpability.”). 119 For example, as Vera Bergelson observes, “[p]ublic views on the allocation of responsibility for rape are well known for their unfairness to the victim,” such that reliance on community sentiment might support “a ‘mini-skirt’ defense to the crime of rape.” Vera Bergelson, Victims and Perpetrators: An Argument for Comparative Liability in Criminal Law, 8 BUFF. CRIM. L. REV. 385, 428–30 (2005). Other empirical work finds the influence of “blatant biases and base sentiments.” Norman J. Finkel, Commonsense Justice, Culpability, and Punishment, 28 Hofstra L. Rev. 669, 702 (2000); see, e.g., Avani Mehta Sood, Attempted Justice: Misunderstanding and Bias in Psychological Constructions of Criminal Attempt, 71 STAN. L. REV. 593, 639 (2019) (“Lay decisionmakers applying the proximity standard in an attempted terrorism case were significantly more likely to construe the thoughts and actions of the defendant as criminal when his name suggested that he was Muslim—even when the legally relevant evidence in the case skewed toward innocence and had nothing to do with Islam. Meanwhile, if the defendant was signaled to be Christian, participants applying the proximity test in an attempted trespass case expressed more positive feelings toward the defendant, rated him as more trustworthy, and perceived him as being more likely to abandon criminal intent.”); HARRY KALVEN, JR. & HANS ZEISEL, THE AMERICAN JURY 164-66 (1966) (postulating a “liberation hypothesis”: that personal sentiments will be more likely to color jurors' judgments when the legally relevant evidence in a case is ambiguous, because “doubts about the evidence free the jury to follow sentiment.”); see also, e.g., Robert J. MacCoun, Biases in the Interpretation and Use of Research Results, 49 ANN. REV. PSYCHOL. 259, 272-73 (1998); Robert J. MacCoun, The Emergence of Extralegal Bias During Jury Deliberation, 17 CRIM. JUST. & BEHAV. 303, 306 (1990); compare Francis X. Shen, Minority Mens Rea: Racial Bias and Criminal Mental States, 68 HASTINGS L.J. 1007, 1030 (2017) (“Do scenario protagonists named Jamal and Lakisha receive more culpable mental state assessments than counterparts named John and Emily? The results suggest that the answer to this question is no.”). 120 Morse, supra note 95, at 514 (that these predispositions may be natural does not mean we should act on them “because, it is commonly held, we cannot derive a normative ought from an empirical is”). Compare GEORGE SHER, IN PRAISE OF BLAME 1 (2005) (criticizing a naturalized account of retribution), with Andrew Oldenquist, An Explanation of Retribution, 85 J. PHIL. 464 (1988) (arguing for a naturalized account of retribution).
C. Why Should We Care About the Guilty Mind?

Why should our blaming practices track the guilty mind? The simplest answer is honesty, which is something that blame lacks when dispensed in disregard of the guilty mind. There are two reasons for this. The first is because of what blame is, namely, a kind of “moral criticism,” often directed toward both the wrongdoer and his or her community.\(^{121}\) The second is because of what blame says, namely, it “expresses a judgment of an actor’s values.”\(^{122}\) As Peter Westen writes:

Morally, blame is indignation on the part of an agent, A, toward another agent, B, for B’s causing a morally wrongful state of affairs, C—indignation being reproach by A toward B for the latter’s motivation in causing C and, specifically, for his causing C out of wrongful disregard for the legitimate interests of others[.].\(^{123}\)

This quote highlights an important point: we perceive the blame meted out by others to reflect the same moral principle (of insufficient concern) upon which our concern for the guilty mind rests. That we would filter third-party blame through our own moral intuitions is not surprising. But it does suggest that blaming practices which disregard the guilty mind would ultimately express—as Sandy Kadish phrases it—“a kind of falsehood,” which is, “to the extent the person is injured by being blamed, unjust to him.”\(^{124}\)

And yet, as the last part of the Kadish quote reveals, this argument is about more than just honesty. Instead, it’s really about blaming in a fair or just way. This concern for fairness or justice is perhaps the most fundamental rationale offered in support of assigning blame upon the guilty mind.

Most often, these fairness-based arguments are framed in terms of desert. For example, one could say (and many have said) that in the absence of a guilty mind, one does not deserve to be blamed, in which case doing so would be unjust.\(^{125}\) But this invites an obvious response: why does deservedness for blame hinge upon the existence of a guilty mind? After all, most people subscribe to the view that the things we do (e.g., the consequences of our actions) are central to determining what we deserve. All else being equal, for example, most of us view those who cause very serious harms (e.g., broken bones) as being more blameworthy than those who commit less serious harms (e.g., hurt feelings).\(^{126}\) In which case it seems plausible that in some situations

\(^{121}\) Kadish, supra note 4, at 264.

\(^{122}\) Westen, supra note 23 at 151; see Mayo Moran, Rethinking the Reasonable Person: An Egalitarian Reconstruction of the Objective Standard. 259-60, 273, 284 (2003).


\(^{124}\) Kadish, supra note 4, at 264; see, e.g., Thomas Scanlon, What We Owe to Each Other 267 (“[[T]he condemnation aspect of punishment is subject to a further requirement: the condemnation must be appropriate. What triggers this requirement is not the unpleasantness of the condemnation, but the content of the judgment expressed.”).

\(^{125}\) See, e.g., sources cited supra note 4; Douglas Husak, Strict Liability, Justice and Proportionality, in the Philosophy of Criminal Law: Selected Essays 152, 154 (2010); Richard Singer & Douglas Husak, Of Innocence and Innocents: The Supreme Court and Mens Rea Since Herbert Packer, 2 Buff. Crim. L. Rev. 859, 860 (1999) (“This concern with whether the conduct of the defendant manifested an evil mind reflects a basic and fundamental principle of justice: Only the blameworthy (guilty), and not the blameless (innocent), [deserve to] be punished.”).

\(^{126}\) See, e.g., Robinson & Darley, supra note 83: Robinson, supra note 83.
where the consequences are particularly egregious (e.g., death), one’s having caused them might be reason enough to deserve being blamed, even in the absence of a guilty mind. Consider, for example, the situation of a driver who, although operating a motor vehicle with extreme caution, kills a small child who darts into the street. Might this individual not deserve at least a little bit of blame?  

The overwhelming position of scholars is “no,” but there are a number of ways to reach that conclusion. One approach is reflected in the honesty-as-fairness argument just noted. Given what blame expresses—namely, a judgment of culpable wrongdoing—a person who lacks a guilty mind does not deserve the false label. Another way to approach the issue, however, embraces a voluntarism-as-fairness argument. The animating idea here is that blame can only be deserved in a situation where the person “could have done otherwise.” This idea asks: how can someone really deserve something—and particularly something unpleasant—for a choice when he or she couldn’t have chosen any differently than he or she did? Whether this (compatibilist) standard of free will is met, in turn, is typically understood to rest upon the existence of things like rationality and self-control, which are the same kind of psychological criteria that serve as preconditions for a guilty mind. Putting all of this together, then, yields the following position: a person who engages in wrongdoing but lacks a guilty mind has not exercised the kind of free will from which we can identify an ability to do otherwise, in which case blame is undeserved—and therefore unjust—under the circumstances.

---

127 See also Scanlon, supra note 10 (“A person who, while doing something he has every reason to believe is quite safe—driving down his street—kills his own child or a neighbor’s child through a freak accident, may understandably find it difficult to live with the fact that he is the one who killed her. [But this] is not a case of moral blame. It is quite natural to reassure such a person by saying that he is blameless.”).
128 See, e.g., Stephen J. Morse, The Moral Methaphysics of Causation and Results, 88 CALIF. L. REV. 879, 879 (2000) (“[V]irtually all criminal law theorists agree that moral fault is at least a necessary condition of blame and punishment[.]”).
129 See sources cited supra notes 121-24.
130 See, e.g., Kadish, supra note 4; HART, supra note 4.
131 See, e.g., Kadish, supra note 4, at 266 (“It may be said in these cases that the person had no effective choice or that no reasonable and upright person could have done otherwise.”); Vincent Chiao, Action and Agency in the Criminal Law, 15 LEGAL THEORY 1 (2009).
132 The compatibilist conception of free will is the weakest of three versions, and the only one which is consistent with a deterministic view of the universe:

Libertarians believe that determinism is incompatible with free will and that the recent literature, therefore, threatens to undermine free will. They argue, however, that humans are free because their decisions are ultimately grounded in irreducibly indeterministic processes, which are likely explained by the principles undergirding the field of quantum mechanics. Hard incompatibilists, on the other hand, believe that determinism is probably true and that both determinism and indeterminism are incompatible with free will. Finally, compatibilists believe that humans possess free will regardless of the truth of causal determinism.

133 See, e.g., Brink, supra note 10; Garvey, supra note 6, at 554.
134 See, e.g., Amy J. Sepinwall, Faultless Guilt: Toward A Relationship-Based Account of Criminal Liability, 54 AM. CRIM. L. REV. 521, 531 (2017) (summarizing this position as: “To impose blame where we could not have done other than what we did, or for the acts of someone whom we cannot control or have no duty to control, is to hold us responsible for something outside of our agency, and so to treat us more harshly than we deserve.”). Consider Sandy Kadish:
Another approach to understanding the relationship among blame, the guilty mind, and fairness focuses less on desert and more on what a reasonable person would do under the circumstances. At the heart of this reasonableness-based approach lies a basic inquiry: what would a person who is just like the actor but who embodies the community’s conception of sufficient concern for others have done if placed in the same situation? This framing effectively focuses our attention on the morally salient factors for which blame is appropriate (namely, objectionable values), while excluding arbitrary influences for which the actor is not responsible (e.g., disability or coercion). This is achieved by imagining the kind of decision (or quality of attention) that someone who possessed the community’s moral values, but the actor’s

[T]he common rationale behind excuses in both law and everyday moral judgments—namely, that justice requires the preclusion of blame where none is deserved. [This rationale would be violated by blaming three different categories of conduct.] In the first category, the person is not to blame because he has no control over his movements; in the second, because though he breached a legal norm, he acted in circumstances so constraining that most people would have done the same; and in the third, because though there is action in breach of a norm in circumstances in which most people would not have done the same, the person, because of a fundamental deficiency of mind, is not a responsible moral agent.

Kadish, supra note 33, at 265.

135 This approach is most often endorsed in the context of evaluating blameworthiness for inadvertent wrongdoing. See, e.g., Westen, supra note 23, at 151; Douglas Husak, Negligence, Belief, Blame and Criminal Liability: The Special Case of Forgetting, 5 CRIM. L. & PHIL. 199, 206 (2011) (“[T]he reasonable person has all of the physical and psychological attributes of the particular defendant with one important exception: the reasonable person has an appropriate degree of concern for others.”). However, it also has purchase in the context of conscious wrongdoing. See, e.g., ALEXANDER & FERZAN, supra note 4, at 147 (“although the question is whether the actor has shown the fortitude necessary to deem him nonculpable, there is some room for individualizing and thus subjectivizing in assessing the difficulty of his choice”); Model Penal Code § 2.02, cmt. at 238 (to determine whether a person’s “conscious disregard of the risk justifies condemnation,” the central question is whether that person’s “disregard of the risk involved a gross deviation from the standards of conduct that a law-abiding person would have observed in the actor’s situation”); id. at 242 n.27 (stating that individualization is appropriate to recklessness as well as negligence); Westen, supra note 23, at 143 (“To assess an actor’s individual blameworthiness by idealized standards that make no allowance for traits over which he has no control is to risk blaming the blameless,” and recommending individualization for intentional/heat-of-passion manslaughter).

136 On Westen’s accounting, a “reasonable person” is an “objective and personified standard” that enables us to evaluate a person’s blameworthiness. Westen, supra note 23, at 151. And because “blame is a judgment of an actor’s values,” this reasonable person must be “a construct that consists of all of what an actor is—including every physical, psychological and emotional trait—except that it also fully embodies [societal] values.” Id.; see Husak, supra note 135, at 206 (“[T]he reasonable person has all of the physical and psychological attributes of the particular defendant with one important exception: the reasonable person has an appropriate degree of concern for others.”).

137 As Westen explains:

Because the test incorporates all of an actor’s physical, emotional and psychological traits, it incorporates all incapacities on his part that are not within his control and not his fault, [and only then] it asks, “How would a person who possessed those incapacities but who otherwise possessed proper values respond cognitively and/or emotionally to the event in question?” The test thus avoids the problem of condemning actors for mental states or emotions they cannot help. At the same time, however, the test also avoids the problem of excusing actors for faults of character they are capable of controlling or compensating for, because while the test incorporates those faults, it expects actors to respond to them in the way that a person with proper values would respond to such dispositions—namely, by subjecting them to the controls and compensatory actions of which he is capable to prevent them from manifesting themselves.

Westen, supra note 23, at 151–52.
shortcomings, would have made (or the kind of attention he or she would have afforded) under the circumstances. If this hypothetical reconstruction indicates that the reasonable person, appropriately individualized, would have behaved no differently, then the actor has done all at that society has reason to expect. In that case, blaming someone who met his or her moral obligations under the circumstances would be unfair.\(^{138}\)

It’s important to highlight that all of these fairness arguments go beyond the injustice wrought by the mere assertion of false blame. They’re equally concerned with the unpleasant consequences that accompany it. In the best case, for example, “[t]he sanctions of personal reproach are gradational and relatively mild, ranging from a raised eye-brow, to verbal chastisement, and to social ostracism[.]”\(^{139}\) However, in a world that is more interconnected than ever, the social ostracism accompanying the assertion of blame can also bring with it life-altering repercussions—for example, the loss of friends, social status, and employment, as well as impairments to one’s mental health and physical well-being.\(^{140}\) Of course, at least some of these consequences may be deserved where the underlying moral responsibility judgment is accurate. However, this additional suffering clearly would not be—and thus would be unjust—in the absence of a guilty mind.

Conditioning blame upon the guilty mind is a norm about fairness, yes, but it may also be a socially beneficial one. Some, for example, understand this norm to be part of a broader network of social ideas and practices that together afford our lives meaning and purpose.\(^{141}\) The idea here

---

\(^{138}\) See, e.g., Kadish, supra note 4, at 266 (arguing that it would be unfair to blame under “circumstances would have led even a reasonable, normally law-abiding person to act in the same way. This being so, his action does not merit blame because it fails to distinguish him from the common run of humankind. It may be said in these cases that the person had no effective choice or that no reasonable and upright person could have done otherwise.”); see id. at 68-69 (“Driving a car, letting a premises, and running a drug packaging business are lawful and socially useful activities . . . Engaging in these activities, therefore, cannot justly serve as a basis for blame. The defendant did only what it was reasonable to do.”); Westen, supra note 23, at 143 (“To assess an actor’s individual blameworthiness by idealized standards that make no allowance for traits over which he has no control is to risk blaming the blameless.”).

It’s worth noting that, in many cases, this injustice will be compounded by the existence of moral hypocrisy. Because it’s quite possible that the person who is unfairly meting out the blame may very well have done the same—or worse still—if placed in the blamee’s situation, which, if true, arguably adds another dimension of unfairness. See Model Penal Code § 2.09, cmt. at 374-75 (the “law is ineffective in the deepest sense, indeed . . . it is hypocritical if it imposes on the actor . . . a standard that . . . judges are not prepared to affirm that they should and could comply with.”); Stupidest Housemaid, J., The Case of the Speluncean Explorers: Revisited I. the Truth, 112 HARV. L. REV. 1917, 1918–19 (1999) (“Regina v. Dudley and Stephens, 14 Q.B.D. 273 (1884), Lord Coleridge, considering a similar case, voted for conviction saying, “We are often compelled to set up standards we cannot reach ourselves, and *1919 to lay down rules which we could not ourselves satisfy.” How very traditional, to support a law with which one has no intention of complying. The stupidest housemaid says “later for that bullshit.”)

\(^{139}\) Westen, supra note 4, at 327.

\(^{140}\) Which is to say nothing of the risk that someone—whether the party doing the blaming or a third party—will resort to extra-legal forms of retribution, such as, for example, the intentional infliction of physical violence.

\(^{141}\) See, e.g., Samuel Pillsbury, The Meaning of Deserved Punishment: An Essay on Choice, Character, and Responsibility, 67 IND. L.J. 719 (1992); Morse, supra note 75, at 62 (“[R]esponsibility practices flow naturally from the criminal law’s traditional view of the person and that mens rea is an essential component of responsibility practices. Moreover, the requirement of mens rea contributes to the meaning and value of our lives as moral beings.”); see also id. at 52-53 (“The capacity for intentional movement and thoughts—the capacity for agency—is a central aspect of personhood and is integral to what it means to be a responsible person. We act because we intend. Responsibility judgments depend on the mental states that produce and accompany our bodily movements. This is how we think
is that by respecting the guilty mind, we reaffirm critical human values (such as personal agency, human dignity, and moral obligation), which in turn ground some of our deepest commitments (for example, personal identity, self-worth, and freedom). Collectively, this network transforms a physical universe comprised of atoms into a moral world comprised of persons. Or at least that’s what those of us think who may already be inclined to believe the “meaning of life” is a topic meriting time and attention. But for those interested in more tangible societal benefits, there’s still much to recommend conditioning blame upon the guilty mind.

First, the general practice may reinforce some useful messages. One is: sufficiently concern yourself with the interests of others—or else. We want to live in a world where people care at least minimally about one another, not only because it’s a nice idea but because such a world will be a safer and more cohesive place to live. So, to the extent that the general practice of conditioning blame upon the guilty mind makes people less likely to act with guilty minds, we all stand to benefit.

Another important message is reflected in the converse: so long as you sufficiently concern yourself with the interests of others, you have little reason to worry about the aversive consequences tied to blame. The argument here is that conditioning blame upon the guilty mind offers those who make reasonable decisions a modicum of security against the prospect of unjustified societal condemnation. This is valuable not only because of how destructive blame can be, but also because most of us can probably point to moments in our lives where the best of intentions were not enough to keep us from bumbling our way into doing the wrong thing. That most of us have a little Steve Urkel inside of us may, for some, be unsettling. But in a world about ourselves, and this is the concept of the person that morality and law both reflect. Law and morality as action-guiding normative systems of rules are useless, and perhaps incoherent, unless one accepts this view of personhood.

---


143 For some good contributions in this area, see, for example, SUSAN WOLF, *MEANING IN LIFE AND WHY IT MATTERS* (2012); ROBERT NOZICK, *The Meaning of Life*, in *PHILOSOPHICAL EXPLORATIONS*, 199-208 (2001).

144 The idea here is that blame “provides occasion for the wider community to enforce or affirm certain norms and values.” Sepinwall, supra note 134, at 537; see, e.g., EMILE DURKHEIM, *THE DIVISION OF LABOUR IN SOCIETY* 60-64 (W.D. Halls transl., Macmillan Press 1984) (1893); David Garland, *Sociological Perspectives on Punishment*, 14 CRIME & JUST. 115, 123 (1991).

145 For further discussion of why this might be the case, see infra notes 1171-73 and accompanying text (discussing empirical desert).

146 As the relevant Wikipedia entry explains:

Steven Quincy Urkel is a fictional character on the ABC/CBS sitcom Family Matters . . .

The character is the epitome of a geek or nerd of the era, due to traits such as large, thick eyeglasses, flood pants held up by suspenders, multi-colored cardigan sweaters, saddle shoes, and a high-pitched voice . . .

From the Urkel character’s debut through the rest of the series’ run, he is central to many of its running gags, primarily property damage and/or personal injury as a result of his inventions going awry or his outright clumsiness. He becomes known by viewers and characters alike for several catchphrases uttered after some humorous misfortune occurs, including “I've fallen and I can’t get up!,” “I don’t have to take this. I'm going home.” “Did I do that?” “Whoa, Mama!” “Look what you
where blame is strictly limited to the guilty mind, we can at least take solace in the fact that so long as we try our best to do right by others, we will be inoculated against some of the most aversive consequences when we fall short.

This, then, is the case in support of conditioning private blame on the guilty mind. It is one primarily rooted in considerations of fairness and societal well-being. To whatever extent these justifications exist in our social lives, though, each is amplified many times over as we turn from the private to the public, and the social to the criminal. This is so due to one key similarity and a few important differences between our social and criminal justice practices.

First, the similarity: the moral judgments rendered by the criminal justice system—namely, conviction and punishment—incontrovertibly constitute a public manifestation of blame. It is widely accepted, for example, that a criminal conviction expresses an official judgment of community condemnation, whereas the formal punishment attached to it—whether probation, one week of jail, one year of prison, or a lifetime of confinement—denotes the extent of that condemnation. We can see this reflected in our propensity to perceive those convicted of a crime to be less than, or morally inferior to, those who have not. Nor is this lowered status

---


147 Morse, supra note 75, at 62–63 (“Which mental states provide what signals is of course a moral and legal normative question that is open to interpretation and revision in light of our best moral and legal theories about culpability, but the importance of mental states is inevitable for creatures such as ourselves. Because we must live interdependently and are always at risk of harm by others, evaluating and responding to the potential harmdoing of others is unavoidable.”).

148 See generally, e.g., Westen, supra note 4, at 326 (“Legal norms of state-imposed punishment differ from personal norms of interpersonal reproach because, even if the two sets of norms have common origins, the institutions and sanctions of state-imposed punishment differ significantly from those of interpersonal reproach.”); Kadish, The Decline of Innocence, 26 CAMBRIDGE L.J. 273 (1968) (“This, of course, is not to embrace the retributive view that responsibility for law violation itself requires punishment, only that responsibility is necessary, but not sufficient, for punishment.”); W.D. Ross, THE RIGHT AND THE GOOD 60-61 (1930); H.L.A. HART, The Presidential Address: Prolegomenon to the Principles of Punishment PROCEEDINGS OF THE ARISTOTELIAN SOCIETY, 9-10 (1959).

149 As Sandy Kadish writes:

[I]ntrinsic to judgments of criminality in our society that they express a moral fault. But this view is surely mistaken. Certainly not all criminal conduct is independently immoral. In some cases the law attaches criminal penalties as well to conduct that, apart from its being prohibited, is not immoral. But in either case criminal conviction charges a moral fault—if not the violation of a moral standard embodied in the criminal prohibition, then the fault of doing what the law has forbidden. The same principle that compels excuses in moral criticism also compels them in the criminal law.

Kadish, supra note 4, at 289.

150 See, e.g., Westen, supra note 4, at 357 (“In reality, the criminal justice system not only officially adjudicates the existence of prohibited conduct but also officially reproaches and condemns the actors whom it finds engaged in it.”); JOEL FEINBERG, DOING AND DESERVING: ESSAYS IN THE THEORY OF RESPONSIBILITY 98 (1970) (“the expression of the community’s condemnation is an essential ingredient in legal punishment”); P. SIMESTER AND G. R. SULLIVAN, CRIMINAL LAW: THEORY AND PRACTICE, 15 (2d ed., 2003) (“When the Court finds an accused guilty of committing a crime, ... there is a public implication that she is blameworthy. ... Paradigmatically, ... censure is inherent in criminal convictions”).
surprising; rather, it’s perfectly consistent with the social meaning of a criminal conviction (and punishment). Consider again Peter Westen, who observes that:

State-imposed condemnation is a public act by which society officially expresses its present moral indignation toward a criminal defendant for violating a criminal prohibition in wrongful disregard of the interests of others.

The point being made here is similar to that previously noted: the social meaning of the criminal law’s formalized judgments of blame is at least roughly consistent with our moral intuitions about the guilty mind. This is true, moreover, in two different ways. The first speaks to threshold liability: to convict someone of a crime is “to adjudge that, rather than being motivated in his conduct by proper regard for interests that the law seeks to safeguard, the person placed insufficient value on those interests.” However, there’s also a second dimension of proportionality, which speaks to the relationship between the extent of an actor’s blameworthiness and the amount of punishment he or she receives.

---

151 As Judge Weinstein observes:

The normal purpose of the criminal law is to condemn and punish conduct that society regards as immoral. Usually the stigma of criminal conviction is not visited upon citizens who are not morally to blame because they did not know they were doing wrong . . . It was inevitable that the development of the criminal law, based as it is upon general and evolving societal mores, would track the development of prevailing views about moral wrongdoing. “The early felonies were roughly the external manifestations of the heinous sins of the day.” . . . The word “felon” itself is a derivative of a Latin term meaning one who is “full of bitterness or venom” and who is “cruel, fierce, wicked, base.

152 Peter Westen, Lex Mitior: Converse of Ex Post Facto and Window into Criminal Desert, 18 NEW CRIM. L. REV. 167, 188 (2015) (“State-imposed condemnation is a moral judgment and, specifically, a moral judgment that the actor disregarded the interests of others. It follows, therefore, that an actor who is condemned for violating a criminal prohibition does not deserve condemnation unless the criminal conduct, in turn, violates interests that are themselves morally legitimate.”); Westen, supra note 4, at 354 (“[B]y publicly declaring an actor to be guilty of a criminal offense, the state expresses indignation at what he has done; and by expressing indignation, the state expresses its belief that he acted with a certain disparaging attitude toward what the criminal statute at hand declares to be the legitimate interests of persons, including himself.”).

153 Westen, supra note 4, at 151; cf. Scanlon, supra note 10 (“Not only blame but also some public expression of blame seems to be called for in cases of mass murder or gross violations of human rights. This is why victims of such crimes, and their families, object so strongly when the perpetrators are not punished and continue to be treated as ordinary members of society. Their claim is not just that it is appropriate to feel indignation and resentment toward such people, or that it is appropriate propriate that they should be made to suffer (although many victims may also want this). The point is also that there is something inappropriate about being asked to treat them as respectable fellow citizens[.]”); see also YAFFE, supra note 4, at 75 (“Criminal punishment necessarily expresses a judgment of criminally culpable wrongdoing; since Cunningham committed no crime, he engaged in no criminal wrongdoing, and was not criminally culpable. Since it’s morally wrong to issue a punishment expressive of such a judgment when the judgment is false, there is a decisive moral objection to criminally punishing Cunningham in this hypothetical.”).
All else being equal, for example, the more deficient the concern manifested by one’s conduct, the more blameworthy we deem the mind behind it to be. The criminal law recognizes this, too, through its willingness to make gradations and scale sanctions accordingly. This is reflected in the law of homicide, which in many jurisdictions would treat:

1. The cold-blooded hitman’s premeditated decision to kill for pay as “first-degree” or “aggravated murder”;
2. The thrill-seeking driver’s spur-of-the-moment decision to speed through a school zone right as the children are being let out as “second-degree” or just regular “murder” (often with informal reference to a “depraved heart” or “extreme indifference”);
3. The hot-blooded parent who decides to shoot the babysitter after walking in on what parent perceives, mistakenly, to be sexual assault of the child as “manslaughter”; and
4. The decision of the blundering college student to get drunk and play with a gun, only to accidently drop it and kill someone in the vicinity as “negligent homicide.”

No doubt, the homicide convictions in each of these cases are similar in that they all express some kind of “negative judgment of the person’s motivating values.” However, the content of these judgments varies in material ways: the degree of the offense, no less than the increasingly severe range of punishment accompanying it, communicates and reflects intuitive differences in blameworthiness.

From an expressive dimension, then, the public blame communicated by the criminal justice system shares many fundamental similarities with the private blame we communicate in our social lives. However, there are also some important distinctions between these two forms of blame, and it is these differences which make the case for conditioning blame upon the guilty mind that much stronger in the criminal justice context.

---

154 See supra Section B.
155 Westen, supra note 23, at 151 (“The attitudes for which persons are blamed range in gravity from maliciousness (e.g., “purpose” to do harm), callousness (e.g., “knowingly” doing harm), indifference to harm, conscious disregard of harm (i.e., “recklessness”), and inadvertent neglect (i.e., “negligence”).”)
156 Westen, supra note 23, at 151.
157 See, e.g., Guyora Binder, The Culpability of Felony Murder, 83 NOTRE DAME L. REV. 965, 1047 (2008); Dan M. Kahan & Martha C. Nussbaum, Two Conceptions of Emotion in Criminal Law, 96 COLUM. L. REV. 269, 352–54 (1996); see also Dressler, supra note 42, at 254 (“Blame, like everything else, is not an all-or-nothing concept. It is one thing to say that people should be held responsible for their actions and quite another to say that they deserve the same punishment as someone else who has committed the same crime.”)
158 A separate but clearly related issue is whether the criminal law blaming norms ought to perfectly track our private blaming norms. Westen, supra note 4, at 326. The position implicit in this paper is that correspondence with private blaming norms is generally a necessary, but not sufficient, condition for justifying a criminal law blaming norm. The differences in consequences between private and criminal law blame offer sound reasons for being far more circumspect about what we criminalize and punish people for than what we might otherwise condemn them for in private. See generally id. (“[E]ven if lawmakers and the public started with identical senses of wrongdoing and blame, one would expect the official rules of criminal law to differ from the ethical rules of interpersonal relationships. Nevertheless, the fact that lawmakers end up making different judgments of wrongdoing and blame in the criminal context than they and the public make interpersonal contexts does not prevent their judgments in law from being judgments of wrongdoing and blame.”).
One important difference is a matter of consequences: however harsh those stemming from private blame can be, on average, the consequences associated with the blame meted out by the criminal justice system are more severe—and sometimes extraordinarily so. This is most obvious as it relates to the consequences for the person charged with a crime. The condemnation and stigmatization inherent in a criminal conviction are typically far greater, and reach a wider audience, than that inherent in private blame. But these are only the informal, social consequences; the formal, legal ones—for example, long-term confinement, death, and post-release collateral consequences—are of unfathomable significance to the individual. And so too is their relative irreversibility: “[I]ndividuals who reproach one another typically know one another and, if they make mistakes, can correct them; while the institutions of official punishment are state officials and random jurors with no personal knowledge of the events and little ability to correct mistakes.”

Another important difference is that these severe consequences are meted out by the state. This matters because the state and its varied public institutions are subject to “normative limits on the ways in which human beings may be treated,” outside of which government action becomes illegitimate. One of those limitations is fairness—that is, the state, if it is to threaten formalized blame and its attendant sanctions, must afford people a fair opportunity to avoid them. But what does this fair opportunity entail? Here, as with the related voluntarism-as-fairness argument discussed earlier, the answer emphasizes the very same things upon which a guilty mind is based—things like rationality, self-control, and choice. But if that’s true, then to the extent the criminal

---

159 Westen, supra note 4, at 327.

160 NICOLA LACEY, STATE PUNISHMENT: POLITICAL PRINCIPLES AND COMMUNITY VALUES 156, 159 (1988); see also Brenner M. Fissell, When Agencies Make Criminal Law, 10 UC IRVINE L. REV. 855, 898 (2020) (“State punishment is a species of coercion, and is thus among the most intrusive forms of state action; even more significantly, though, this coercion takes the form of violence.”).

161 See, e.g., HART, supra note 4, at 181 (“[U]nless a man has the capacity and a fair opportunity or chance to adjust his behaviour to the law its penalties ought not be applied to him.”); Peter Cane, Responsibility and Fault: A Relational and Functional Approach to Responsibility, in RELATING TO RESPONSIBILITY: ESSAYS FOR TONY HONORÉ ON HIS EIGHTIETH BIRTHDAY 81, 108 (Peter Cane & John Gardner eds., 2001) (“As agents, we have an interest in freedom of action, in being able to act without incurring the serious penalties and blame that attach to criminal responsibility.”).

162 As Amy J. Sepinwall observes:

The fault principle (along with the voluntary act requirement) ensures that individuals will not be punished for acts they could not avoid—those that the agent does not will (e.g., those performed while she is unconscious), or performed in innocent ignorance of the risk they impose, or under circumstances that compel or coerce the agent’s criminal conduct. In this way, individuals can protect themselves from the threat of state punishment, because they can be assured that the circumstances warranting punishment are within their control.

Sepinwall, supra note 134, at 530–31; see, e.g., LACEY, supra note 160, at 146 (“Both rationality and the capacity for responsible action are thus for liberalism at once factual features of human nature and sources of normative limits on the ways in which human beings may be treated, particularly by political and other public institutions.”); Brink, supra note 10; Garvey, supra note 6, at 554; HART, supra note 4, at 152 (emphasizing “those who ‘had, when they acted, the normal capacities, physical and mental, for doing what the law requires and abstaining from what it forbids, and a fair opportunity to exercise these capacities’.”)
law’s judgments of blame disregard the guilty mind, they constitute illegitimate exercises of state power.\footnote{See, e.g., Garvey, \textit{ supra} note 6, at 547 (“The state gets to decide what makes for a guilty mind when it defines the mental-state elements of a crime, and indeed it should get to decide, as long as it has the authority to do so. But any authority is subject to limits, and the guilty mind as mens rea is one such limit. No state can legitimately punish an actor unless he committed a crime with mens rea.”); HART, \textit{ supra} note 6.}

These (among other) differences suggest that the fairness and societal welfare-based arguments which support respecting the guilty mind in the social context apply with even greater force in the criminal justice context. For example, if principles of honesty, desert, and reciprocity indicate it would be unjust to blame someone who lacks a guilty mind in private, then it would be that much greater of an injustice for the criminal justice system to do so publicly, where the consequences are both life-altering and beyond the state’s legitimate authority to impose.\footnote{See also Model Penal Code § 2.09, cmt. at 374 (“[L]aw is ineffective in the deepest sense, indeed . . . it is hypocritical, if it imposes on the actor who has the misfortune to confront a dilemmatic choice, a standard that his judges are not prepared to affirm that they should and could comply with if their turn to face the problem should arise. Condemnation in such a case is bound to be an ineffective threat; what is, however, more significant is that it is divorced from any moral base and is unjust.”); cf. Westen, \textit{ supra} note 4 (“Interpersonal reproof is just that: it is personal and typically private or semi-private; while official condemnation is purposefully impersonal and purposefully public.”).}

A similar point can be made about societal welfare. To the extent conditioning private blame upon the guilty mind supports the background conditions for a meaningful life, reinforces community values,\footnote{See also Model Penal Code § 2.09, cmt. at 374 (“legal norms and sanctions operate . . . in the fashioning of values and of character); Meir Dan-Cohen, \textit{Responsibility and the Boundaries of the Self}, 105 HARV. L. REV. 959, 1001–02 (1992) (“[W]e commonly expect the law to comport with our ordinary notions of responsibility. The law is expected to reinforce people’s sense of responsibility by making an explicit public pronouncement regarding a certain instance of responsibility and by dramatizing its significance through severe sanctions. The legal recognition of my authorship of a certain object or event is supposed to strengthen my identification with the appropriate responsibility base.”); Morse, \textit{ supra} note 75, at 59–60.}

and creates a sense of security from unjustified intrusions, then all the more so is this true in the public context given (again) the broader audience for, greater forms of suffering inherent in, and more irreversible nature of public blame.\footnote{Consider Henry Hart:}

For these reasons, the already compelling case for conditioning blame upon the guilty mind in the private context is that much stronger in the context of the criminal justice system. But that

---

\textit{Henry M. Hart, Jr., The Aims of the Criminal Law, 23 LAW & CONTEMP. PROBS. 401, 410 (1958).}
does not mean that it is decisive. Because in both contexts, there are also costs that flow from conditioning blame in this way.\textsuperscript{167} Here, I will focus on the costs for the criminal justice system by examining the traditional public safety-based argument against utilizing mens rea in assigning public blame; however, much of what I will say will ring true about private blame, too.

The conventional legal argument against guilty minds is typically articulated in terms of the public safety benefits that accrue from holding people strictly liable for their wrongdoing. To make sense of this argument, it’s helpful to unpack two key concepts underlying it: (1) strict liability and (2) public safety benefits.

There are a number of ways the criminal law might (and does) hold people strictly liable.\textsuperscript{168} One way is at the level of an offense. For example, a pure strict liability offense of assault of a police officer (APO) would punish all instances in which one person causes bodily injury to a police officer, without regard to whether the person acted purposely, knowingly, recklessly, or negligently as to any aspect of the offense. In that case, a person who trips on the sidewalk and stumbles into a police officer could be convicted of APO, although his accident was reasonable and the officer was completely outside of the person’s field of vision (in which case the person had absolutely no reason to know who the victim was).

Another way the criminal law might hold someone strictly liable is at the level of an element. For example, a partial strict liability version of APO might require proof of either purpose, knowledge, recklessness, or negligence with respect to causing bodily injury to someone, but no culpable mental state whatsoever as to whether the victim was a police officer. In that case, the previously mentioned klutz could not be convicted, but someone who purposely punched an undercover, plainclothes officer (say, in a bar fight), with absolutely no reason or way to know who the victim was could be convicted.

These two forms of strict liability emphasize the narrow, PKRN understanding of mens rea. However, yet another version focuses on the criminal law’s failure to recognize the moral salience of even broader aspects of a guilty mind. This would include, most obviously, holding someone strictly liable in the sense of denying them the ability to raise an affirmative defense—for example, insanity and duress—which negates blameworthiness.\textsuperscript{169} But it also arguably speaks to the recognition of broader impairments of moral judgment, which—as illustrated by Part I’s analysis of the APO convictions in the Crossland case—might go beyond the scope of current

\textsuperscript{167} See, e.g. Kadish, supra note 4, at 271 (“While justice is a fundamental value, and hence usually paramount, it may on occasion be outweighed by these other values. What is right is what is right all things considered.”).


\textsuperscript{169} Consider, for example, the situation of a person who purposely punches a police officer, but is (1) suffering from mental illness so severe that he is unable to recognize the immorality of the conduct or (2) was coerced to do so in the face of a credible, violent assaultive threat by a third party.
excuse defenses yet still call into question whether someone’s state of mind is sufficiently blameworthy to support a criminal conviction under the circumstances.

That’s a rough overview of the varieties of strict liability. Now let’s consider another important component of the conventional legal argument against guilty minds: how strict liability is understood to promote public safety.

Typically, proponents of this argument emphasize a couple of different policy ideas. The first sounds in deterrence; it holds that policies and doctrines of strict liability incentivize people to be more careful—and more generally desist from criminal activity—in the future. The second idea sounds in incapacitation; it holds that someone who engages in wrongdoing may still be dangerous even if she is not blameworthy, in which case incapacitating this individual will stop her from committing further crimes in the future.

These two intuitively appealing policy ideas have led many to conclude that, all else being equal, mens rea requirements must surely be worse for public safety than a strict liability alternative given the comparative lack of deterrence (i.e., weaker message about the risks of wrongdoing) and incapacitation (i.e., passing up opportunities to take dangerous people off the streets). In that case, limiting the criminal law to the guilty mind can only be justified by our commitment to fairness—and that simply may not be worth the price.

This standard accounting of mens rea, strict liability, and public safety surely seems plausible. But that doesn’t mean it actually is. That’s because, among other things, these policy ideas rest upon empirical assertions about the relationship among how we structure criminal sanctions, how humans respond to them, and the consequences of imposing them. And as a number of scholars and reports have highlighted, we have good reason to question them. A full exploration of those reasons is beyond the scope of this paper; however, the most important points can be summarized thusly.

First, for deterrence to work, the potential offender must: (1) know of the rule and sanction; (2) rationally calculate the personal cost to himself of violating the rule, discounted by the likelihood of detection, and weigh that against the benefits of engaging in criminal action; and (3) be willing and able to conform his or her conduct to that calculus. How often are these requirements satisfied? Not often, it appears. Which explains why there’s little empirical support for the idea that specific doctrinal manipulations have any meaningful impact on criminal behavior insofar as the decision making of individual criminal actors is concerned.

Second, achieving effective incapacitation through the manipulation of individual criminal law rules rests upon similarly dubious grounds. In one sense, of course, all incapacitation is by definition effective. For the duration of the time that a person is locked up, she can’t commit crimes on the outside. However, she can commit crimes on the inside, and some prisoners do, which detracts from the wellbeing of other prisoners. But even more fundamental is the fact that most people we imprison will someday be released. Yet prison has a documented criminogenic effect: after lengthy stays, some will be more likely to commit crimes than they were when they were incarcerated. And just as problematic is the fact that we’re quite bad at predicting who will and will not commit crimes—indeed, only moderately better than pure chance. What’s more, any predictive prowess we do possess is largely rooted in immutable characteristics—things like gender, age, and employment history—which are at best morally irrelevant to blame and at worst discriminatory grounds for blaming.

Third, deterrence and incapacitation are but two trees in a broader forest of public safety, and narrowly focusing on them—as politicians have been apt to do—could risk burning down the entire forest. That’s because of the underappreciated public safety costs that flow from employing criminal sanctions in a manner that conflicts with people’s perceptions of fairness. The general idea here, which Paul Robinson has labeled “empirical desert,” theorizes that distributing criminal liability and punishment in accordance with people’s shared intuitions of justice is a critical part of controlling crime for three main reasons:

(1) The criminal law depends upon voluntary compliance and cooperation.

(2) Whether people do in fact comply and cooperate with the criminal law depends, at least in part, upon whether they view the law as legitimate.

(3) Whether people view the criminal law as legitimate depends upon—among other things—whether the imposition of liability and scaling of punishment tracks people’s shared intuitions of justice.

The theory of empirical desert is relevant to an evaluation of the alleged public safety benefits of strict liability because a great many justice intuitions we possess relate to the guilty mind, which is something the criminal law might or might not track. And as Paul Robinson has illustrated through decades of empirical work, the numerous strict liability exceptions to the criminal law’s general presumption of mens rea seem to fall into the “might not” category. But if that’s true, then it poses a problem for the traditional assumption that strict liability is an effective means of promoting public safety. Because in situations where strict liability policies lead to judgments that conflict with our intuitions about the guilty mind, the results may be criminogenic: a greater sense of alienation from the law, a decrease in respect for the law, and, ultimately, more frequent violations of the law by the citizenry. Where, in contrast, the criminal law is aligned

---


172 Paul H. Robinson, Strict Liability’s Criminogenic Effect, 12 CRIM. L. & PHIL. 411, 426 (2018); see, e.g., sources cited supra note 170.

173 As Robinson observes:
with community sentiment on the guilty mind, the theory of empirical desert presents the opposite
effect: enhanced legitimacy, greater levels of social trust, more voluntary compliance, and, ultimately, increased public safety.\textsuperscript{174}

None of which is to say that the theory of empirical desert offers a decisive public safety-based argument against \emph{ever} holding wrongdoers strictly liable—or even that the theory resoundingly supports the general (if inconsistent) commitment to mens rea reflected in American criminal law. For there are numerous other factors that further complicate any attempt to understand the relationship between strict liability and public safety. Consider, for example, just a few possibilities:

(1) defendants whom the community believes to possess a guilty mind will use the evidentiary challenges associated with establishing mens rea requirements to avoid criminal liability, in which case mens rea policies will lead to outcomes that are perceived as unjust and lead to greater disillusionment with the law; or

(2) prosecutors and factfinders will use the increased discretion afforded to them by a strict liability regime to prejudicially convict some groups, but not others, thereby producing outcomes that are racially disparate and lead to greater disillusionment with the law.

(3) the resources spent on prosecuting and incarcerating those who lack a guilty mind would more effectively control crime if invested elsewhere (i.e., opportunity costs).

In the final analysis, then, respect for the guilty mind might come at the cost of public safety, or it might not. One thing we can reasonably be certain of, however, is that we really don’t know. This is so, I would argue, because in most situations we simply lack the empirical knowledge and modes of computation and prediction that one would need to reliably assess whether a strict liability regime would actually promote public safety better than would an approach that closely tracks the guilty mind, all things considered. If that’s true, however, it brings with it important consequences. Because we can also be reasonably certain of the fact that a principled commitment to conditioning blame upon the guilty mind will produce a more \emph{just} society. And in a system of that defines itself by that metric, that’s probably as good a reason for the criminal justice system to respect the guilty mind as one can hope for.

\textsuperscript{174} See, e.g., sources cited \textsuperscript{supra} note 170.

As the criminal law’s moral credibility is incrementally reduced, the system is incrementally more likely to inspire resistance and subversion rather than acquiescence and cooperation. Witnesses are less likely to report crimes and help investigators. Jurors are less likely to follow their legal instructions and more likely to substitute their own intuitions. Police and prosecutors are less likely to follow the system’s rules and more likely to morally justify their subversions of those rules. Citizens are less likely to defer to the criminal law in the grey areas of criminality, and less likely to worry about a stigmatizing effect from criminal conviction. Most importantly, people are less likely to defer to and internalize the criminal law’s norms. Criminal law can harness the powerful forces of stigmatization, social influence, and internalized norms only if it has established itself as a moral authority. And the use of strict liability and the criminalization of regulatory violations can only undermine that goal.

Robinson, \textit{supra} note 172, at 426.