Decoding Guilty Minds: How Jurors Attribute Knowledge and Guilt

Matthew R. Ginther

Francis X. Shen

Richard J. Bonnie

Morris B. Hoffman

Owen D. Jones

See next page for additional authors

Recommended Citation

Decoding Guilty Minds: How Jurors Attribute Knowledge and Guilt

Authors

Matthew R. Ginther, Francis X. Shen, Richard J. Bonnie, Morris B. Hoffman, Owen D. Jones, and Kenneth W. Simons

ESSAY

Decoding Guilty Minds: How Jurors Attribute Knowledge and Guilt

Matthew R. Ginther[a]
Francis X. Shen[β1]
Richard J. Bonnie[γ]
Morris B. Hoffman[δ]
Owen D. Jones[ε]
Kenneth W. Simons[ζ2]

A central tenet of Anglo-American penal law is that in order for an actor to be found criminally liable, a proscribed act must be accompanied by a guilty mind. While it is easy to understand the

---

importance of this principle in theory, in practice it requires jurors and judges to decide what a person was thinking months or years earlier at the time of the alleged offense, either about the results of his conduct or about some elemental fact (such as whether the briefcase he is carrying contains drugs). Despite the central importance of this task in the administration of criminal justice, there has been very little research investigating how people go about making these decisions, and how these decisions relate to their intuitions about culpability. Understanding the cognitive mechanisms that govern this task is important for the law, not only to explore the possibility of systemic biases and errors in attributions of culpability but also to probe the intuitions that underlie them.

In a set of six exploratory studies reported here, we examine the way in which individuals infer others' legally relevant mental states about elemental facts, using the framework established over fifty years ago by the Model Penal Code ("MPC"). The widely adopted MPC framework delineates and defines the four now-familiar culpable mental states: purpose, knowledge, recklessness, and negligence. Our studies reveal that with little to no training, jury-eligible Americans can apply the MPC framework in a manner that is largely congruent with the basic assumptions of the MPC's mental state hierarchy. However, our results also indicate that subjects' intuitions about the level of culpability warranting criminal punishment diverge significantly from prevailing legal practice; subjects tend to regard recklessness as a sufficient basis for punishment under circumstances where the legislatures and courts tend to require knowledge.

INTRODUCTION

All across America, every day, we ask jurors to decide what some stranger, who stands accused of criminal behavior in the past, was thinking at the time of his alleged crime. That's hard to do. And despite the fact that the task is ubiquitous, and crucial to the government's legitimacy in depriving citizens of physical liberty, the ways that jurors actually impute mental states to defendants is almost entirely unknown. If jurors aren't operating the way the legal system assumes they are, then gross injustices—wrongful convictions and wrongful acquittals—may be pervasive.

Cognitive science has provided some clues. We have learned some of the cognitive mechanisms at play when a person attempts to infer another's state of mind and intentions.[3] And these "theory of mind" studies have even begun to uncover the neural foundations for taking the perspective of another.[4] Yet the law's interest, more specifically, is in how potential jurors attribute mental states to defendants while operating under established legal frameworks. Little empirical research has addressed that issue.[5]

---

3. *See generally* IAN APPERLY, MINDREADERS: THE COGNITIVE BASIS OF "THEORY OF MIND" (2011).

4. *See, e.g.,* R. Saxe et al., *Understanding Other Minds: Linking Developmental Psychology and Functional Neuroimaging,* 55 ANN. REV. PSYCHOL. 87 (2004).

5. *See* Matthew R. Ginther, Francis X. Shen, Richard J. Bonnie, Morris B. Hoffman, Owen D. Jones, René Marois & Kenneth W. Simons, *The Language of Mens Rea,* 67 VAND. L. REV. 1327, 1331–34 (2014) (summarizing the extant literature).

The legal frameworks are quite specific. As most readers will recall, statutes typically require both a bad act (actus reus) and a culpable mental state (mens rea) for criminal liability to attach. Further, most states follow the Model Penal Code's ("MPC") approach of using a hierarchy of culpable mental states (which correspondingly calibrate the crime committed and the punishment earned). Although state implementations vary somewhat, the MPC's mental state hierarchy includes these: purposeful, knowing, reckless, and negligent.[6]

A 2011 article published by several of us cut initial inroads into how jury-eligible subjects applied the MPC framework to mens rea determinations. In that study—*Sorting Guilty Minds*[7]—we provided subjects with short written criminal scenarios in which an interior sentence provided a transparent window into the mental state of a harm-causing protagonist. For instance, when the actor was reckless, we told subjects that: "The offender was aware of a substantial risk that his actions would cause the victim's death." We then asked subjects, in various interconnected experiments, either to sort these different mental states by MPC categories or instead to simply choose appropriate punishment amounts.

Those experiments revealed two troubling things. First, although subjects could appreciate the differences in language that communicated a blameless actor compared to one acting with purpose to cause a harmful result, subjects had much more trouble distinguishing between negligent, reckless, and knowing actors. Second, the punishments subjects imposed suggested they do not see any consistent moral distinction (which law often assumes) between reckless and knowing actors.[8]

As with any groundbreaking research, *Sorting Guilty Minds* raised as many questions as answers. Two of those unanswered questions are central to the new research and results reported here. First, how do subjects impute mental states when they are not given a transparent window into the thoughts of defendants? Second, how do subjects map mental states, as they have been defined by the MPC, onto judgments about criminal culpability?

The six new experiments reported here, generously funded by the MacArthur Foundation Research Network on Law and

---

6. MODEL PENAL CODE § 2.02 (AM. LAW INST. 1980).

7. Francis X. Shen, Morris B. Hoffman, Owen D. Jones, Joshua D. Greene & René Marois, *Sorting Guilty Minds*, 86 N.Y.U. L. REV. 1306 (2011).

8. Our 2014 article confirmed and extended these results by demonstrating that the difficulties subjects experienced were not simply semantic. For even when the language used to define and communicate mental states was clarified, the reckless and knowing mental states remained relatively indistinguishable to subjects. Ginther et al., *supra* note 5, at 1360.

Neuroscience,[9] provide the first rigorously empirical answers to those questions. Each experiment sheds important new light on how jury-eligible subjects actually perform the mind-reading tasks our criminal justice system assigns to them.

For example, our studies reveal that, with little to no training, jury-eligible Americans can apply the MPC framework in a manner that is largely congruent with the basic assumptions of the MPC's mental state hierarchy. At the same time, however, our results also provide cause for concern. For instance, subjects' intuitions about the level of culpability warranting criminal punishment diverge significantly from prevailing legal assumptions and practice, because subjects tend to regard recklessness as a sufficient basis for punishment even under circumstances in which the legislatures and courts tend to require knowledge.

We proceed in three Parts. Part I provides background and context, and introduces the common elements and basic design of our experiments. Part II presents the key findings. Part III discusses their implications—including for deepening our understanding of juror behavior, for further informing potential efforts at reform, and for future research.

I. EXPLORING MENTAL STATE ATTRIBUTION

To be convicted, one must not only have engaged in prohibited conduct but must ordinarily have done so with a guilty mind. The requirement of establishing a guilty mind, commonly referred to as mens rea,[10] is a predicate of modern criminal statutes[11] and has

---

9. *See generally* MACARTHUR FOUND. RES. NETWORK ON L. & NEUROSCIENCE, http://www.lawneuro.org (last visited Oct. 19, 2017) [https://perma.cc/6H7C-UL7U].

10. The term mens rea is adopted from the longstanding legal axiom of *actus reus non facit reum nisi mens sit rea*, or "the act is not culpable unless the mind is guilty." This famous phrase dates back at least approximately one thousand years. Francis Bowes Sayre, *Mens Rea*, 45 HARV. L. REV. 974, 983 (1932).

11. This is perhaps no more evident than in modern drug laws. A recent survey indicates that forty-eight of the fifty states require that a defendant both have possession of the drugs and *know* he is possessing illegal drugs as an element of the offense. The two exceptions are Florida and Washington, where state legislatures have eliminated this knowledge provision in favor of an affirmative defense. *See* State v. Adkins, 96 So. 3d 412, 423 n.1 (Fla. 2012) (Pariente, J., concurring) (citing a survey done in Dawkins v. State, 313 Md. 638, 646–49 (1988), though correcting for North Dakota revising the law to include a willfulness provision and Florida's decision in *Adkins* upholding the constitutionality of a statute without a knowledge requirement); State v. Bradshaw, 98 P.3d 1190 (Wash. 2004). Even then, in both Florida and Washington defendants charged under the statute are provided the opportunity to raise the affirmative defense that they did not have knowledge of the illicit nature of the substance. *Id.* ("[T]he defendant can concede all elements of the offense but still coherently raise the 'separate issue' . . . of whether the defendant lacked knowledge of the illicit nature of the controlled substance." (citation omitted)); City of Kennewick v. Day, 11 P.3d 304, 310 (Wash. 2000) (noting the affirmative defense is

constitutional significance, as illustrated by recent Supreme Court jurisprudence.[12] And the requisite mens rea must be established separately for each element of the offense. Broadly construed, these elements are categorized as pertaining either to the defendant's conduct, to the result of the conduct, or to the attendant circumstances.

    In previous work we explored how well subjects could distinguish different mens rea language as it pertained to the result elements of an offense. This is important, for instance, because in many states a jury finding of a purposeful (P) or knowing (K) killing could yield decades in prison, while a finding that the same killing was reckless (R) or negligent (N), in the alternative, could result in as little as probation.[13] And yet we found our subjects struggled to differentiate the language associated with knowing and reckless conduct.

    The importance of differentiating mental states is not limited to result elements. It is often, in fact, more crucial for circumstance elements.

    For result elements, the significance that mens rea plays in grading homicide offenses is largely focused on whether the defendant purposefully, knowingly, or merely recklessly caused the victim's death. However, there are few other instances where mens rea as to a result plays as critical a role.

    With circumstance elements, by contrast, disputes about mens rea as to attendant circumstance elements are quite frequently at the crux of culpability determinations. In most (though not all[14]) statutes

---

available by "one of two alternative showings: (1) that the defendant did not know he was in possession of the controlled substance . . .; or (2) that the defendant did not know the nature of the substance he possessed" (citation omitted)).

    12.   Staples v. United States, 511 U.S. 600, 619 (1994) (holding that "[s]ilence does not suggest that Congress dispensed with *mens rea* for the element . . . at issue"); United States v. X-Citement Video, Inc., 513 U.S. 64, 69, 77–78 (1994) (declining to enforce the plain meaning of the statute criminalizing distribution of child pornography and instead creating a knowledge requirement with respect to the elements of the age of the minor performers and the sexually explicit nature of the material).

    13.   *See* Shen et. al., *supra* note 7, at 1348 n.98 (noting that "[m]ost states use either the K/R distinction as an express distinction in the definition of levels of homicide, or effectively do so by making K an aggravator to an R homicide").

    14.   Even if the defendant did not *know* that a material circumstance existed, some criminal statutes permit a criminal conviction if he or she was aware of a "substantial risk" that it did ("recklessness") or if he or she should have been aware of such a risk under an objective standard ("negligence"). In rare cases, criminal liability may be "strict" because no mens rea is required for the material circumstance. The typical case is statutory rape where even a non-negligent belief that the underage party was over the statutory age would not be exculpatory. For its part, the MPC strongly discourages the use of strict liability for all but minor violations but does make a single exception in the case of statutory rape of a child under the age of ten. MODEL PENAL CODE §§ 213.4(4), 213.6(1) (AM. LAW INST. 1980). For a more thorough discussion of strict liability offenses, see generally Kenneth W. Simons, *When is Strict Criminal Liability Just?*, 87 J. CRIM. L. & CRIMINOLOGY 1075 (1997).

in which a material circumstance is part of the offense, *knowledge* of the inculpatory circumstance is the threshold requirement for criminal liability. For example, even if a defendant transports hazardous waste without a permit, he's not generally guilty of any crime at all unless he *knows* the waste was hazardous. He might have been negligent in not knowing, or even reckless in taking the chance. But for this crime, as for many circumstance crimes, *knowledge* is required for liability to attach. This same analysis differentiates innocent and guilty conduct across a number of different offenses. For example, did he *know* there were drugs in his car? Did he *know* the kid he just sold alcohol to was underage? Did he *know* that the tax return didn't include all his income? Did he *know* that the car he received was stolen? Did he *know* the chemicals were going to be used to make methamphetamine? Did he *know* that his emails contained classified material? In this way, compared to decoding guilty minds for result elements, decoding guilty minds as they pertain to circumstance elements is both a much more ubiquitous and consequential task in assessing culpability under a criminal statute. For this reason, the present study focuses its examination solely on circumstance elements.

Given its importance, defining the contours of what exactly constitutes a "guilty mind" has been a central issue in criminal law throughout history.[15] Modern legislation usually follows the Model Penal Code's four-part approach to categorizing mental states.[16] With respect to circumstance elements, the relevant mental states are defined as follows by the MPC.[17]

> ➢ Knowingly. A person acts knowingly with respect to a circumstance when he is aware that the circumstance exists.

> ➢ Recklessly. A person acts recklessly with respect to a circumstance when he consciously disregards a substantial and unjustifiable risk that the circumstance exists. The risk must be of such a nature and degree that, considering the nature and purpose of the actor's conduct and the circumstances known to him, its disregard involves a gross deviation

---

15. Sayre, *supra* note 10, at 988.

16. For an exposition of the historical development, see Paul H. Robinson, *A Brief History of Distinctions in Criminal Culpability*, 31 HASTINGS L.J. 815 (1980).

17. *See* MODEL PENAL CODE § 2.02 (AM. LAW INST. 1980). We note that while "purposefulness" is defined by the MPC, it practically never applies for circumstance elements. Consideration of the defendant's purposefulness regarding the existence of a *circumstance* arises only in relation to inchoate offenses.

from the standard of conduct that a law-abiding person would observe in the actor's situation.

➤ Negligently. A person acts negligently with respect to a circumstance when he should be aware of a substantial and unjustifiable risk that the circumstance exists. The risk must be of such a nature and degree that the actor's failure to perceive it, considering the nature and purpose of his conduct and the circumstances known to him, involves a gross deviation from the standard of care that a reasonable person would observe in the actor's situation.

➤ Blamelessly. A person acts blamelessly when he does not have any of the culpable mental states defined above.

Variations of this language appear in a majority of jurisdictions in the United States.[18] Even jurisdictions that have not legislatively endorsed the MPC's approach often employ its mental state categories.[19] The Supreme Court has frequently done so in recent years when interpreting federal statutes that lack an explicit mens rea requirement.[20]

Given the MPC's wide adoption and influence, our empirical focus in this study is on how everyday subjects understand and apply the MPC's culpability hierarchy when decoding the mental state of others. To achieve this, we employed six interrelated experiments. In all of them, jury-eligible subjects read a series of short scenarios online and then answered a question. (Details on the subjects appear in Appendix A.) Each scenario presented a series of facts about a fictional protagonist named "John." The scenarios were designed to enable lay subjects to characterize John's mental state. We used a "fact pattern–scenario" paradigm in which we crafted nine distinct (or "core") fact patterns, each of which we could alter slightly in order to communicate evidence compatible with attributions of different mental states.[21]

---

18.   *See* Paul H. Robinson & Jane A. Grall, *Element Analysis in Defining Criminal Liability: The Model Penal Code and Beyond*, 35 STAN. L. REV. 681, 691–92 (1983).

19.   JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 33 (4th ed. 2006) ("[C]ourts, on their own, sometimes turn to the Model [Penal] Code and its supporting commentaries for guidance.").

20.   An example is the 2016 decision in *Voisine v. United States*, 136 S. Ct. 2272 (2016). In analyzing the scope of the statutory language "misdemeanor crime of domestic violence" involving the "use . . . of physical force," the Court relied entirely on MPC mental state categories. *Id.* at 2276. It held that reckless assaults, no less than knowing or intentional assaults, were embraced by this language. *Id.* at 2282.

21.   This was similar to the approach taken in Ginther et. al., *supra* note 5, and Shen et. al., *supra* note 7.

"Scenario," as used here, and as illustrated momentarily, refers to an individual, mental-state-specific version of the core fact pattern.

Each of the nine core fact patterns involved a different criminal offense. Each offense used in our experiments has an attendant circumstance element. We ensured that our fact patterns spanned a spectrum of wrongful conduct, ranging from selling alcohol to a minor to harboring a fugitive, and we used a preliminary study to validate that we were successful in this regard.[22] Table 1 lists the nine core fact-pattern offenses used in our experiments.

TABLE 1: THE CORE FACT PATTERNS

Basic Fact Pattern	*Attendant Circumstance*
Drug Trafficking: John is accused of driving a car over the border with drugs in the trunk.	Did John know that the drugs were in his trunk?
Theft by Receiving: John is accused of buying goods that were stolen.	Did John know that the goods were stolen?
Sale of Alcohol to Underage Person: John is accused of selling alcohol to an underage person.	Did John know that the person was underage?
Statutory Rape: John is accused of having sex with an underage person.	Did John know that the person was underage?
Tattoo of a Minor: John is accused of giving a tattoo to a minor.	Did John know that the person was a minor?
Illegal Hiring: John is accused of hiring a person not authorized to work in the United States.	Did John know that the person was not authorized to work in the United States?
Harboring a Fugitive: John is accused of harboring a fugitive.	Did John know that the individual was a fugitive?
Insurance Fraud: John is accused of filing a false claim.	Did John know that the submitted claim contained false information?
Unlawful Carrying of Loaded Firearm: John is accused of carrying a loaded firearm in public.	Did John know that the firearm was loaded?

Unbeknownst to subjects, we generated four different scenarios from each of the nine core fact patterns. Specifically, each of the core fact patterns yielded one scenario each for the knowledge, recklessness, negligence, and blameless mental states.[23] For instance, there were four different versions of the drug trafficking fact pattern, and each of these four varied only in the mental state dimension. In Table 2 we show how this worked for a single, illustrative scenario.

---

22. Data on file with author and available upon request.

23. Again, the mens rea of purposefulness almost never applies to circumstance elements. *See supra* note 17.

Each fact pattern was a paragraph in length, ranging from 80 to 160 words. The complete set of fact patterns is provided in Appendix B. For context, each fact pattern also included a simplified version of the relevant criminal statute.[24] For instance, the drug trafficking fact pattern began with the statement: "A person is guilty of unlawful trafficking of illegal drugs if the person transports an illegal drug."[25]

Each fact pattern then proceeded with one or two introductory sentences that set up the facts relevant for that fact pattern. For example, the drug trafficking fact pattern always started with:

> "John is an American citizen who regularly travels to Nogales, a Mexican town near the U.S. border where John has many relatives, including an uncle. Several of John's relatives have told John that his uncle hangs out with local drug dealers."

These opening facts were the same in each of the knowledge, recklessness, negligence, and blameless scenarios.

Following the opening facts, the second component communicated information bearing on John's mental state. As an example, the blameless scenario version of the fact pattern provided:

> "When he travels, John makes every effort to avoid interacting with his uncle, but one night despite all the precautions John takes, when John is asleep the uncle secretly breaks into John's car and hides a large package of cocaine inside the spare tire compartment, where it is not visible to the naked eye."

The final sentence of the scenario completed the facts, making clear that the circumstance in question was present. This final sentence was also, like the opening facts, identical across each of the four mental state scenarios generated for each fact pattern. For example, in the drug trafficking fact pattern, all four scenarios ended with the text:

> "When John tries to re-enter the United States, the border patrol agents who inspect John's car at the border

---

24. These statutes were not meant to reflect the language of any specific jurisdiction.

25. Except where otherwise noted in specific experiments below, all language pertaining to the required mental state for culpability was excised from the statute. This was because we were not interested in testing subjects' ability to interpret the statute, but rather the facts of the scenario themselves.

open up the package given to John by his uncle and find
that it contains cocaine."

Constructed in this way, our fact patterns resembled the typical
circumstance case in the real world: there is no doubt that the
circumstance exists (e.g., the drugs were in the car), but there is
potentially much doubt about whether the actor *knew* those drugs were
in the car or was *aware* of the possibility that the drugs were in the car.

The key aspect of this experimental design is that we
manipulated the surrounding circumstances to strengthen or weaken
the inferences that might be drawn about John's awareness of the
material facts, while controlling other aspects of the fact pattern. This
manner of manipulating the scenarios reflects the realities of criminal
trials wherein the mental state of another person at a particular time
(specifically whether or not that person was aware of a particular fact
or suspected it) is inferred from their behavior and from the
surrounding circumstances.

We recognize that in the case of recklessness, negligence, and
blamelessness, the application of these concepts also includes value
judgments about whether the person (John) should have refrained from
acting as he did in our scenarios. Recognizing this, we constructed our
scenarios to vary the circumstances to "best fit" the distinctions drawn
by the MPC drafters between knowledge, recklessness, negligence, and
blamelessness. Specifically, we constructed four kinds of scenarios
designed to convey different fits, which we have labeled "manifest
knowledge," "awareness of risk," "negligently unaware," and
"blameless."

The "manifest knowledge" scenarios include objective evidence
about John's own conduct tending to show that he knows, as in this
example, that drugs are in the trunk of the car (e.g., he opens the duffel
bag containing the drugs, looks inside it, and puts it in the trunk). These
are meant to elicit the conclusion that John had actual knowledge of the
circumstance in question. Because knowledge is directly demonstrated
through the facts of the scenario, we regard all responses other than
knowledge to be erroneous.

The "awareness of risk" scenarios, by contrast, only include
inferential evidence about the circumstances. This inferential evidence
includes John's observations of the behavior of other people. For
example, in the same fact pattern as above, John's uncle, who associates
with drug dealers, pays him one hundred dollars to bring a duffel bag
across the border. As part of the scenario we tell subjects that "[h]is
sister says 'I don't think that's a birthday present . . .' but his uncle says
(with a wink towards John) that 'it's just coffee.' " Based on the totality

of the circumstances, we think the best fit for these scenarios is recklessness, but knowledge and even just negligence could also be plausible responses, depending on a subject's judgment about the strength of the appropriate inferences.

The "negligently unaware" scenarios include evidence that would make some people suspicious but wouldn't necessarily make everyone suspicious, and was thus intended to elicit negligence as the best fit. Using the same example, subjects are told that "John's uncle gives John a giftwrapped package to bring to a friend, and when John asks what's in the box, John's uncle says it is a birthday present for the friend." For these scenarios, we define the best fit as negligence, but recklessness and blamelessness are plausible responses.

Finally, the "blamelessness" scenarios include evidence that the person takes active and reasonable precautions to avoid criminal conduct and thus is not criminally culpable. Again using the same example, "[w]hen he travels, John makes every effort to avoid interacting with his uncle, but one night despite all the precautions John takes, when John is asleep the uncle secretly breaks into John's car and hides a large package of cocaine inside the spare tire compartment, where it is not visible to the naked eye." Based on the circumstances presented, we regard all responses aside from blameless to be erroneous.

For all experiments, we randomized the order in which subjects encountered the scenarios and mental states. Subjects read one and only one scenario from each of the nine fact patterns. After reading each scenario, subjects were asked to answer a question, which varied (as described in the next Section).

TABLE 2: ILLUSTRATION – VARYING MENTAL STATE WITHIN A SINGLE FACT PATTERN

Part 1 of 3

John is an American citizen who regularly travels to Nogales, a Mexican town near the U.S. border where John has many relatives, including an uncle. Several of John's relatives have told John that his uncle hangs out with local drug dealers.

Mental State	Part 2 of 3	Part 3 of 3
Manifest Knowledge	One night when John is in Nogales, John's uncle asks him to bring back a large duffel bag with a "birthday gift" to a friend in the U.S. His sister says, "I don't think that's a birthday present . . . " but his uncle says (with a wink towards John) that "it's just coffee." John agrees to transport the package for $100, opens the duffel bag before he puts it in his trunk, and nods when he sees that the bag contains cocaine.	
Aware of Risk	One night when John is in Nogales, John's uncle asks him to bring back a large duffel bag with a "birthday gift" to a friend in the U.S. His sister says, "I don't think that's a birthday present . . . " but his uncle says (with a wink towards John) that "it's just coffee." John agrees to transport the duffel bag for $100 and puts it in his trunk without opening it.	When John tries to reenter the United States, the border patrol agents who inspect John's car at the border open up the package given to John by his uncle and find that it contains cocaine.
Negligently Unaware of Risk	One night when John is in Nogales, John's uncle gives John a giftwrapped package to bring to a friend, and when John asks what's in the box, John's uncle says it is a birthday present for the friend. John agrees to deliver the box, and doesn't open it up to look inside.	
Blameless	When he travels, John makes every effort to avoid interacting with his uncle, but one night despite all the precautions John takes, when John is asleep the uncle secretly breaks into John's car and hides a large package of cocaine inside the spare tire compartment, where it is not visible to the naked eye.	

II. THE SIX EXPERIMENTS

A. *Experiment 1: Attributing Mental States Using MPC Definitions*

The MPC has established a hierarchy of mental states relating to the circumstance elements of an offense. Experiment 1 set out to answer a straightforward question: Do subjects reliably apply this hierarchy as the drafters assumed they would? We presented subjects with the scenarios, as described above, and then asked them to select the wrongdoer's mental state, with a prompt specific to the mental state in question. For example, in the case of the drug trafficking scenario, subjects were asked:

> "Which of the following mental states best describes John's mental state as to the fact that drugs are in the bag?"

Following that prompt, subjects were provided with the four mental states and their definitions and asked to select which best represented the mental state of John as to the pertinent circumstance.[26]

The results from Experiment 1, summarized in Table 3, demonstrate that the typical juror grasps the hierarchy established by the Model Penal Code. The leftmost vertical column indicates what we refer to as the best fit mental state. Moving rightward, each column identifies in turn the mental state actually selected by subjects for those scenarios. More specifically, the cells with figures indicate the percentage of subjects who selected a given mental state described in the top row, for each scenario type described in the leftmost column. For instance, 6% of subjects attributed a negligent state of mind to protagonists in scenarios intended to describe manifest knowledge. The gray-shaded diagonal of cells, from the upper left to the lower right, identifies the percentage of subjects that identified the best fit for the protagonist's mental state.

---

26. The definitions were the MPC definitions, save for two slight modifications. First, because the MPC definitions were written to apply not only to circumstance elements but result and conduct elements as well, we excised the results- and conduct-based components from each of the definitions we provided (with minor rephrasing so it worked as a stand-alone sentence). Second, we customized each definition so it referred to the specific facts in each scenario. For instance, we changed the definition of "knowingly" from "[a] person acts knowingly with respect to the attendant circumstances when he is aware that such circumstances exist" to "[a] person acts knowingly with respect to the duffel bag containing illegal drugs when he is aware that the drugs are in the bag."

TABLE 3: JURY-ELIGIBLE AMERICANS CAN APPLY THE MPC FRAMEWORK IN A MANNER THAT IS LARGELY CONGRUENT WITH THE BASIC ASSUMPTIONS OF THE MPC'S MENTAL STATE HIERARCHY

Scenarios	*Subject Chose:*			
	Knowing	Reckless	Negligent	Blameless
Manifest Knowledge	81%	12%	6%	1%
Aware of Risk	28%	40%	30%	2%
Negligently Unaware of Risk	8%	28%	53%	11%
Blameless	3%	4%	20%	74%

As expected, we observed high correspondence between subjects' responses and our best fit response for the manifest knowledge and blamelessness scenarios—81% and 74%, respectively. However, for our "aware of the risk" and "negligently unaware of the risk" scenarios, there was the possibility of multiple plausible responses, and concordantly the correspondence between subjects' responses and what we labelled the best fit was commensurately lower—40% and 53%, respectively. In addition to correspondence between responses and the best fit, we examined how individuals tended to err in their responses. In other words, did individuals tend to err towards more or less culpable determinations? For "aware of the risk" scenarios, subjects were about as likely to select negligence (30%) as knowledge (28%). However, for the "negligently unaware of the risk scenarios," subjects strongly erred towards ascribing awareness of the risk to the offender, with 28% selecting recklessness, while only 11% selected blamelessness.

A graphic depiction of the subjects' performance in applying the MPC definitions to the scenarios is displayed in Figure 1, below. The black shading represents the best fit between the MPC mental states and the evidence described in scenarios, while the lightest gray shading represents clearly erroneous responses. Dark gray reflects responses that were plausible but inclined towards greater culpability, while mid-gray reflects responses that were plausible but inclined towards less culpability. Generally speaking, the subjects tended toward the best fit, but they also varied in their attributions.[27]

---

27. One interesting empirical question is whether individual subjects show state- or trait-based differences in how they ascribe mental states. Our data is not well-suited to answer this

FIGURE 1: APPLYING MPC MENTAL STATE DEFINITIONS

■ Best Fit   ▦ Plausible - More Culpable

▦ Plausible - Less Culpable ▦ Erroneous

MANIFEST KNOWLEDGE
81%   19%

AWARE OF RISK
40%   28%   30%   2%

NEGLIGENTLY UNAWARE OF RISK
53%   36%   11%

BLAMELESS
74%   26%
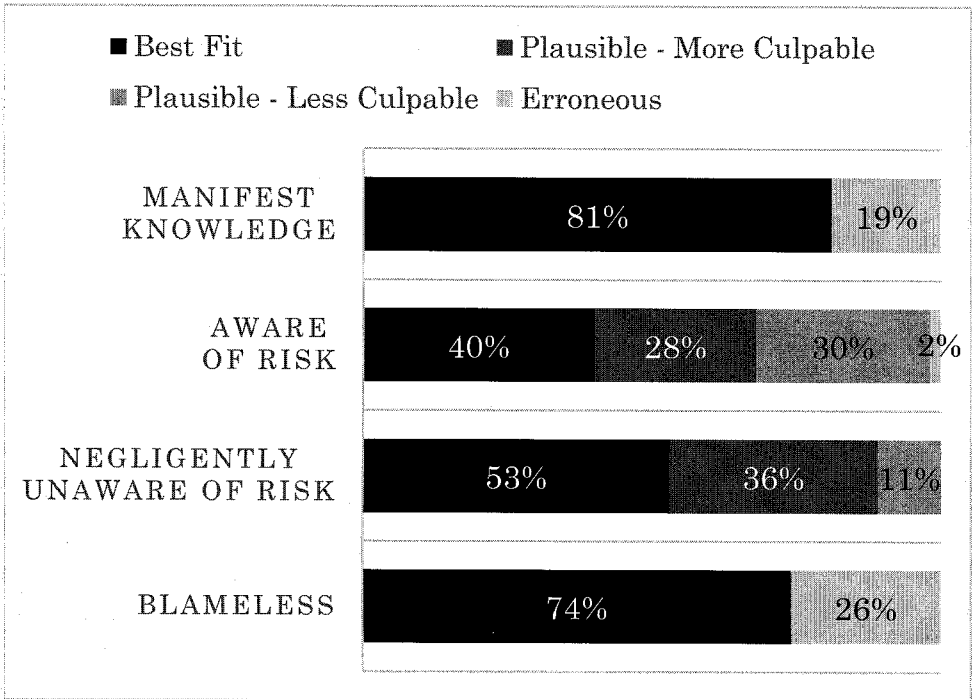
B. *Experiment 2: Subjects' Unguided Moral Intuitions About What Mental State Is Sufficient for Criminal Liability*

What level of culpability do individuals regard as sufficient for punishment? Experiment 2 was designed to ascertain how ordinary jury-eligible persons would answer this question when given free rein to do so: we asked them to decide whether or not they would convict John *without giving them a mens rea instruction*. In other words, after evaluating a scenario, they were asked to apply a statute that contained only an actus reus, with no mention of any requisite mens rea for the attendant circumstance. By asking whether they would find the defendant guilty under the statute, this question served as a proxy for asking the subjects whether John's level of culpability, as demonstrated in the particular scenario (i.e., our best fit scenarios for K, R, or N) provided a sufficient basis for punishment.

---

question, though further research could evaluate the extent to which this may account for some of the intersubject variability.

For example, subjects evaluating the drug trafficking scenario received the following prompt:

> Criminal Statute for Unlawful Trafficking: A person is guilty of unlawful trafficking of illegal drugs if the person transports an illegal drug.
>
> Based on the evidence provided do you think that John is guilty of unlawful trafficking?

This question is similar to the types of questions that psychologists ask when they use what are called "detection and discrimination" tasks. Such tasks help to answer the question: What type of information is necessary for an individual to draw a particular conclusion? In the present case, this detection and discrimination analysis applies because we are seeking to discover the level of culpability (as defined by the MPC) that is sufficient for subjects to return a guilty verdict under a statute that lacks any specific mens rea requirement.

Figure 2 displays a "decision curve" that plots the percentage of guilt determinations on the y-axis (vertical axis) and the mental state on the x-axis (horizontal axis). This kind of plot allows us to examine whether the shape of the decision curve presents a discrete "inflection point"—that is, a disproportionate jump in percentages—as our subjects move from sorting scenarios into one mental state to another. If so, we call this point the "decision threshold."[28] The heavily shaded line presents the average of all nine scenarios while the light dashed lines present the average for each of the individual scenarios.

---

28. What we refer to here as a decision curve is equivalent to what psychologists refer to as a psychometric function. The inflection points of these functions (where the slope is the highest or where the second derivative is equal to zero) is a commonly used indicator of a sensory or decision threshold in not only psychology, but signal detection more generally. For more information see generally Neil A. Macmillan, *Signal Detection Theory*, *in* 4 STEVENS' HANDBOOK OF EXPERIMENTAL PSYCHOLOGY (Hal Pashler ed., 3d ed. 2002) (using inflection points as an indication of the location of a decision or sensory threshold).

FIGURE 2: THE RELATIONSHIP BETWEEN BEST FIT MENTAL
STATE AND PERCEIVED CULPABILITY

Mental State of Offender
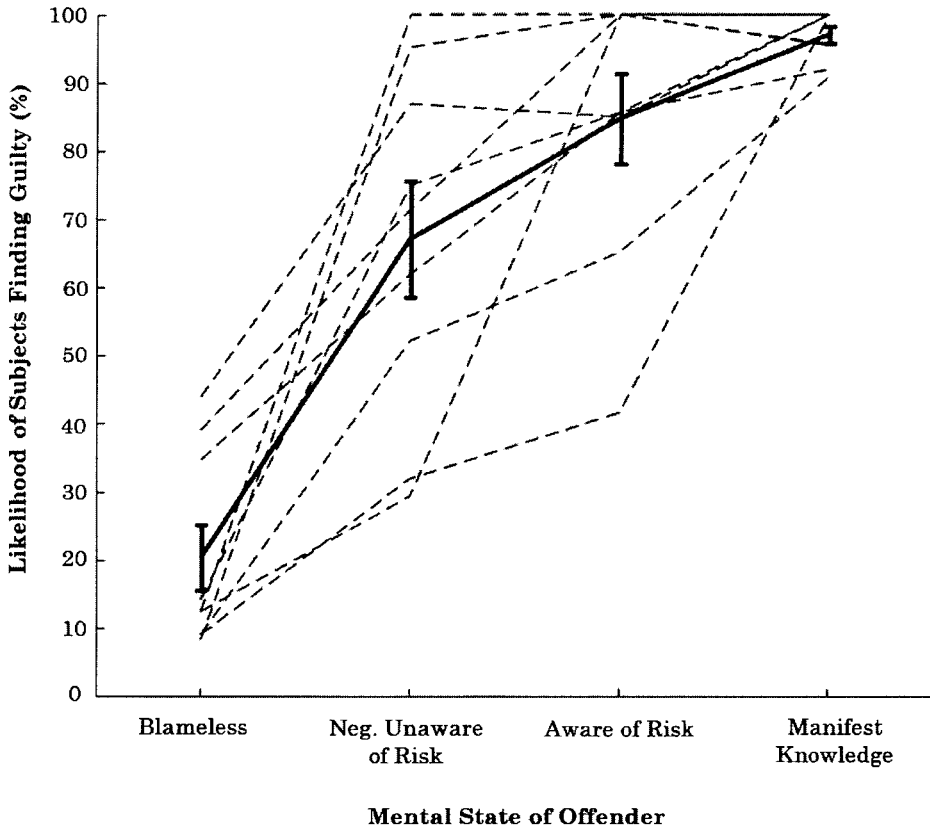
Our first finding in Experiment 2 is that there is substantial variability across the nine scenarios, particularly for the "negligently unaware of the risk" and "aware of the risk" categories.[29] This is, however, consistent with the study design (which allowed a range of plausible attributions of these two mental states) and with the finding in Experiment 1 that the subjects varied substantially in their selections when exposed to scenarios describing protagonists with these two mental states. A second finding is that for all but two of the

---

29. A natural question would be: What are the characteristics of the scenarios that lead to more culpability as opposed to less? While this is an interesting question and we provide the full tables in the appendix, the study was not designed to make claims about individual scenario fact patterns, and so we omit discussion of individual scenarios here since any conclusions would be purely speculative. We do note, however, that linking scenario characteristics with intuitions about culpability is an area of research we intend to explore.

statutory contexts, four out of five subjects exposed to strong evidence that the material circumstance was present (our best fit scenario for recklessness) regarded this evidence as a sufficient basis for punishment. Moreover, we found that two out of three subjects found that weak, but suspicious, evidence of the material circumstance (our best fit for negligence) was regarded as sufficient for criminal punishment under the statute. The shape of the decision curve supports the conclusion that the decision threshold for culpability exists at negligence, with much more modest incremental increases in the percentage finding guilt above the level of negligence.

In summary, when unguided by any instruction regarding the mens rea required for conviction, most jury-eligible subjects find an offender guilty even when the evidence of knowledge is murky. Why are subjects so willing to convict on the basis of evidence that only raises a suspicion that the actor was aware of the material circumstance? One potential explanation lies in the results of Experiment 1, which demonstrated that when evaluating such scenarios, individuals have a bias towards attributing awareness when the evidence is ambiguous. For example, in the scenarios where the best fit response was negligence, three-fourths of the subjects who did not select negligence selected knowledge or recklessness as opposed to blamelessness. This finding suggests that jury-eligible adults tend to be willing to attribute knowledge or awareness of risk (suspicion) to hypothetical defendants based on equivocal evidence. Another possible explanation is that jury-eligible adults find recklessness, and in many cases negligence, to be a morally sufficient basis for criminal liability in the context of the "circumstance" offenses explored in this study. In the sequence of experiments beginning with Experiment 3 we attempt to disentangle the subjects' attributions of mental state from their moral judgments about liability.

C. Experiment 3: Use of Signals to Align Subjects' Responses with the Best Fit Mental State

In order to make more precise connections between mental state and culpability, we adopt from our previous two sets of published experiments[30] the use of "signals" to convey information about John's mental state. For instance, to communicate that John was reckless in the drug trafficking fact pattern, we told subjects, "John thinks there is a good chance that the duffel bag contains illegal drugs." The text of all signals is provided in Table 4.

---

30. *See* Ginther et. al., *supra* note 5; Shen et. al., *supra* note 7.

TABLE 4: SIGNALING LANGUAGE

Knowledge
John realizes [that the circumstance exists].
John knows [that the circumstance exists].
Recklessness
John thinks there is a good chance [that the circumstance exists].
John thinks it is quite possible [that the circumstance exists].
Negligence
It doesn't occur to John [that the circumstance might exist].
John does not suspect that [that the circumstance might exist].
John is distracted and does not consider [that the circumstance might exist].
John gives no thought to the possibility [that the circumstance might exist].

We used signals for K and R to push the "awareness of the risk" scenarios towards either knowledge or recklessness, while we used the negligence signals to buttress the fact that the "negligently unaware of the risk scenarios" did, in fact, describe negligent conduct. So that subjects did not think that the absence of a signal in the "manifest knowledge" scenarios was indicative of anything, we also included knowledge signals in those scenarios as well.

True, real jurors don't get direct access, like this, to a defendant's prior thoughts. But the value of using explicit signals nonetheless, in the present experiment, is to discover the relationships between the mental states of offenders, on one hand, and subjects' intuitions about culpability, on the other.

Experiment 3 repeated Experiment 1, with the sole difference being the use of the "signaled" scenarios. The results from Experiment 3 are displayed in Table 5. As might be expected, the signaling language had no significant impact on the responses to unambiguous scenarios reflecting manifest knowledge and blamelessness, but it did nudge subjects' responses toward the best fit mental state attributions for the scenarios meant to reflect knowledge, recklessness, and negligence. In general, the signaling produces results more in line with the best fit mental state that was signaled in the scenario.

Specifically, when using the knowledge and recklessness signals we were able to shift "awareness of risk" scenario responses—which had previously been relatively evenly distributed among K/R/N (28%/40%/30%) in Experiment 1—towards either K (which increased from 28% to 67%) or R (from 40% to 58%), respectively. Similarly, the responses to the "negligently unaware of the risk" scenarios shifted toward negligence (from 53% to 63%), but more importantly, subjects no longer displayed a robust predisposition to err towards awareness of

the risk when the negligence signals were used (compare responses selected for N scenarios in Table 3 with Table 5).

One question posed by these results is why the effect of the signal was not even more substantial in aligning the answers with the best fit mental state. The most likely explanation, in our view, is that the information communicated in the scenarios is fairly complex, including (1) several sentences about the surrounding circumstances that were designed to raise different levels of suspicion about the presence of the material fact, and (2) a simple declarative statement by an omniscient observer about John's mental state. We suspect that the complexity of the stimulus weakened the impact of the mental state signal—which would have been determinative if it had been given by itself without any additional evidence (because it would have amounted to nothing more than a matching exercise). Nonetheless, the increased accuracy and the removal of strong error biases in the case of negligence provided us with greater precision with which to make conclusions about how the typical juror assesses the culpability of an actor with these prescribed mental states.

TABLE 5: APPLYING MPC MENTAL STATE DEFINITIONS WHEN MENTAL STATE IS EXPLICITLY SIGNALED

Scenarios	Subject Chose:			
	Knowing	Reckless	Negligent	Blameless
Manifest Knowledge	76%	16%	6%	1%
Aware of Risk Signaled as Knowledge	67%	21%	12%	0%
Aware of Risk Signaled as Recklessness	18%	58%	24%	0%
Negligently Unaware of Risk Signaled as Negligence	1%	20%	63%	16%
Blameless	2%	5%	16%	78%

D. Experiment 4: Subjects' Unguided Intuitions About Culpability
Sufficient for Criminal Liability: Revisited

In Experiment 3 we repeated the procedures of Experiment 1, but using signals to nudge subjects towards the best fit mental state in each scenario. Having established in Experiment 3 that this was largely successful, in Experiment 4 we repeated the procedures of Experiment 2, except again using signals to clarify subjects' intuitions about the link between mental state and culpability. Again, we included statutory definitions of the offenses applicable to each fact pattern that did not include a mens rea requirement.

Our data, plotted in Figure 3, reveal a key new finding: First, as compared with Figure 2, the variability among scenario fact patterns was significantly reduced, as intended. But most significantly, the data show that for most people, the decision threshold of culpability lies between negligence and recklessness. To see this from the Figure, notice that there are nine dashed lines, each of which shows the subjects' decision curve for one of the nine given scenarios. The dark, unbroken line is the average of those nine. The inflection point is recognizable at the negligence-recklessness boundary, which reflects the largest (and therefore steepest) jump up, from the prior mental state, in the percentage of subjects willing to convict. Indeed, as Figure 3 depicts, the percentage willing to convict jumped from under 50% for those exposed to the scenarios signaled as negligence to over 80% of those exposed to the scenarios signaled as recklessness.

FIGURE 3: RECKLESSNESS IS A THRESHOLD FOR CULPABILITY

Mental State of Offender
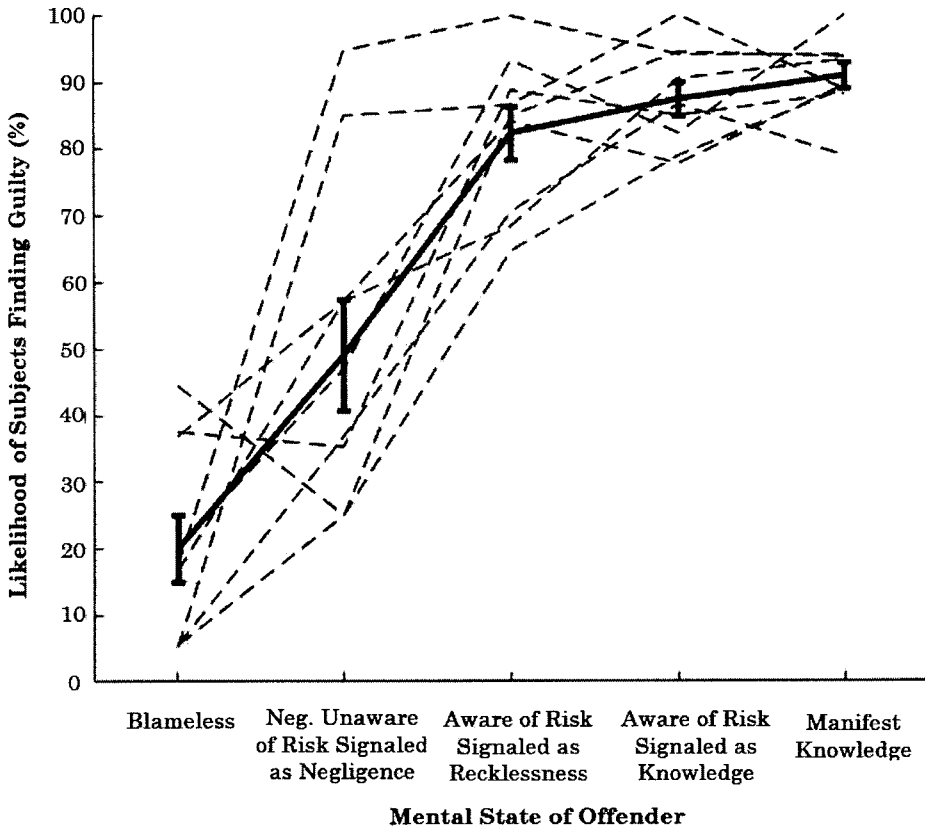
Note that there is very little difference between the proportion of people finding an offender guilty if he had a mens rea of recklessness (82%) compared to a mens rea of knowledge (88%), even if it was manifested by conduct (91%). Put another way, for a large proportion of the subjects, recklessness was both necessary and sufficient for conviction.

To be clear, this doesn't mean that people find perpetrators with knowledge to be only a few percentage points more blameworthy than people performing the same act recklessly. And it doesn't mean that people somehow think that those with knowledge should only be punished a few percentage points above the punishments that the reckless receive. What it does suggest is that roughly 80% of the population—if interpreting a statute without any mention of the mens rea requirement as to a relevant circumstance (that is, unconstrained by mental state categories, of the MPC or otherwise)—appears to

believe that the mental state the law tends to call "reckless" is fully sufficient to warrant a guilty verdict and conviction.

Of course, this finding immediately suggests the tantalizing question to investigate next: Would instructing the subjects about the mens rea required for conviction by the applicable statutes make any difference?

E. Experiment 5: The Effect of Instructions

Experiment 4 showed that most subjects required recklessness (awareness that the circumstance might exist) before they were willing to find John criminally liable, but did not require knowledge that the circumstance actually existed. We designed Experiment 5 to assess the subjects' response to an instruction requiring knowledge as the mens rea before criminal liability would attach. Put another way, would expressly informing subjects that the statute actually required knowledge push the culpability threshold up from recklessness to knowledge? Experiment 5 thus had to differ in two ways from Experiment 4.

First, at the beginning of Experiment 5 we provided subjects with an instruction page similar to the guidance provided by a judge to a jury. This page instructed subjects on the basics of element analysis and how bad acts need to be accompanied by a culpable mental state in order to hold an individual criminally responsible.[31] We also provided subjects with slightly modified versions of the MPC's definitions of knowledge, recklessness, and negligence.[32] And we instructed them to carefully note the mental state demands of the statute.

Second, we altered each statute to expressly require knowledge as an element. This addition, alone, appeared bolded in the statute presented to subjects. For example, for the drug trafficking fact pattern, subjects were told:

> "A person is guilty of unlawful trafficking of illegal drugs
> if the person **knowingly** transports an illegal drug."

---

31. The instruction read as follows:

A crime is committed when the defendant has performed an illegal act **accompanied** by a culpable mental state. That means proof of the commission of the act alone is not sufficient to prove that a crime has been committed. The defendant must also have a culpable mental state to certain attendant circumstances such as someone's age. Culpable mental state means either knowingly, recklessly or negligently, as described below.

32. *See supra* text accompanying note 14.

The bolded term—"knowingly"—was defined immediately below the statute and in a manner specific to the facts of the case.

Despite these explicit directions, the results from Experiment 5 indicate that subjects did *not* measurably alter their decision threshold for culpability. Figure 4 illustrates the finding by comparing the decision curve from the last experiment—Experiment 4 (the "No Instruction" condition)—with the new decision curve, reflecting the "Knowledge Instruction." What's immediately striking, of course, is that the curves are virtually identical. Specifically, the percentage of subjects finding a defendant signaled as reckless guilty of violating a statute that explicitly required actual *knowledge* was virtually the same as the percentage who held the defendant liable when no mens rea was prescribed—82% and 81%, respectively.

What these results demonstrate is that, at least in laboratory conditions, juror intuitions about whether the defendant who behaved negligently or recklessly when engaging in the illegal conduct should be convicted appear to be unaffected by an explicit instruction prescribing knowledge as the mental state required for conviction.

FIGURE 4: COMPARISON OF INSTRUCTION AND NO INSTRUCTION
CONDITIONS
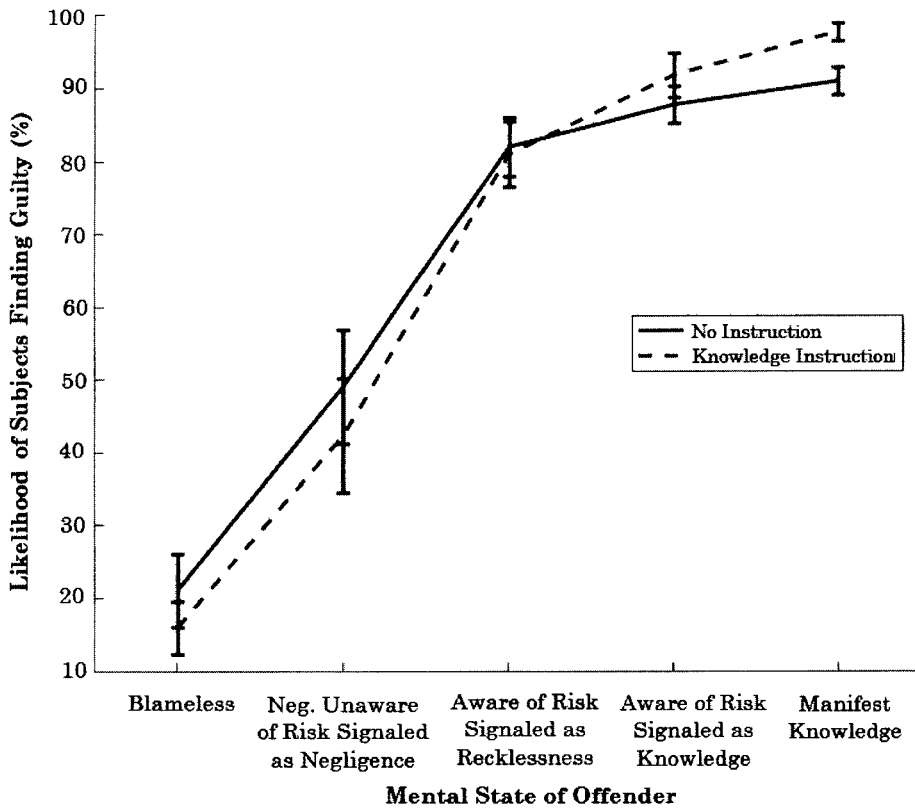
F. Experiment 6: The Effect of Instructions: Revisited

A puzzle emerges when comparing the results of Experiments 3 and 5. Recall that in Experiment 3 about 60% of the scenarios signaled as demonstrating awareness of risk were characterized by the subjects as constituting recklessness under the definitions provided (see Table 5). However, in Experiment 5 we saw that more than 80% of subjects evaluating these same scenarios believed John was guilty of violating a statute that conspicuously required *knowledge* as a predicate for conviction (see Figure 3). Can these two findings be explained or reconciled? Two possibilities come to mind: Perhaps twice as many subjects in Experiment 5 found the distinction between knowledge and recklessness to be confusing (when no definitions were provided) as did those in Experiment 3 (when definitions were provided). Alternatively,

perhaps being told that knowledge was required for guilt affected the subjects' attributions of mental state.

To investigate further, we conducted Experiment 6, which was identical to Experiment 5, except for the addition of a new question. After presentation of the scenario and after eliciting the subjects' decisions about whether the defendant should be convicted, subjects were also asked to make a specific finding about the defendant's mental state—i.e., to identify and select the defendant's mental state from the list of definitions specifically tailored to the facts of the scenario. The results, depicted in Figure 5, show that in the scenarios signaled as involving awareness of risk (what we have been calling "recklessness" scenarios), subjects were more likely to attribute knowledge to the actor when told that knowledge was required for conviction (33%) than when they were told to apply the definitions without being told about the legal consequences of their finding (18%). Thus, Experiment 6 seems to show that knowing that "knowledge" was required for conviction made subjects more likely to attribute knowledge to the actor in ambiguous scenarios than they would otherwise have been in a context where they were *not* told the legal consequences of the attribution.

FIGURE 5: COMPARISON OF HOW SUBJECTS CLASSIFY AWARE OF THE RISK SCENARIOS BASED ON WHETHER THE APPLICABLE STATUTE HAD A KNOWLEDGE INSTRUCTION

■ Knowing ■ Reckless ■ Negligent ■ Blameless

AWARE OF RISK WITH K INSTRUCTION:

33% 38% 27% 2%

AWARE OF RISK WITH NO INSTRUCTION:

18% 58% 24%
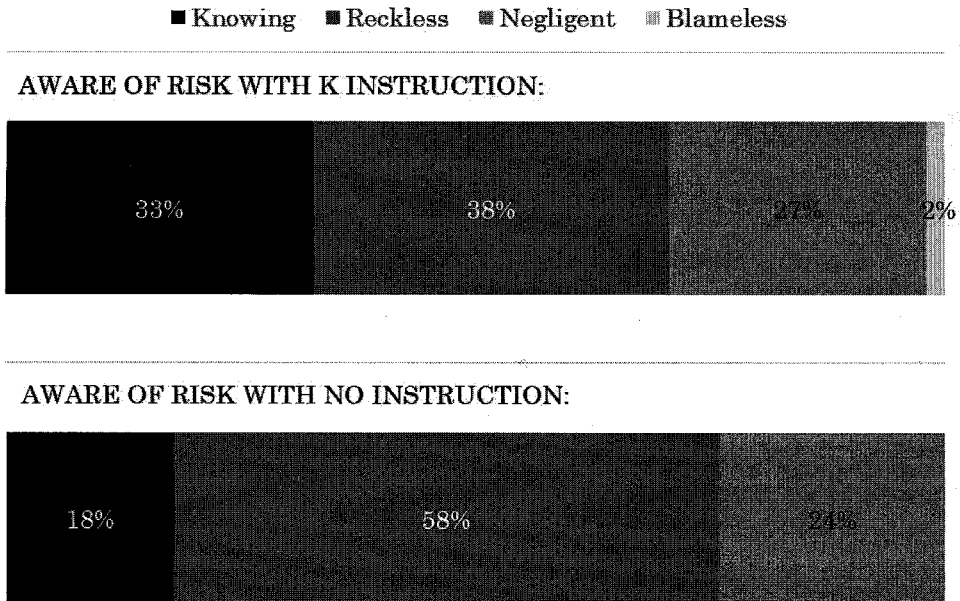
III. DISCUSSION

These experiments were designed to study attribution of mental states based on postulated observations of a criminal actor's conduct and the circumstances clearly known to him. They built on the foundation laid in earlier studies by the research team, which used structured scenarios to vary the harm caused by the actor and the actor's mental state regarding that harm. Mental states were operationalized based on the MPC's hierarchical framework. The present study differed from those studies in two major respects. First, rather than focusing on mens rea as to result elements (e.g., the death of the victim), this study focused on the actor's mental state regarding a material circumstance (e.g., the age of a sexual partner) in the definition of the offense that differentiates lawful conduct from criminal conduct. Second, whereas the previous study stipulated the actor's mental state, the current study was designed to elicit subjects' attributions of mental state to the actor based on descriptions of the objectively observable circumstances relating to the actor's conduct.

A. Attributing Mental States

A man is stopped at the U.S. Border and his duffel bag contains a large container of cocaine. In the United States it is illegal to "knowingly or intentionally . . . possess a controlled substance."[33] The man claims he had no idea the drugs were in the bag. In the absence of a confession or direct observation of the defendant's incriminating conduct, how are jurors to decide what the defendant knew or believed? Do they do it in an acceptable manner? In this exploratory study, we used scenarios to represent varying levels of evidence regarding the existence of the legally material circumstance—such as the presence of cocaine in a bag—as a basis for exploring the subjects' attributions of a putative defendant's mental state pertaining to that circumstance.

On the upper end of the culpability hierarchy, "knowing" that the circumstance exists (K) is distinguished categorically from all cases where it is suspected but not "known" to exist (R). In order to isolate clear cases of K, we modeled a case where the actor manifests "knowledge" by his own conduct such as acknowledging the existence of the material circumstance; in all other cases, however, whether the actor "knows" or "believes" the circumstance exists or only "suspects" that it does is a probabilistic attribution based on the strength of the evidence.

---

33.   21 U.S.C. § 844 (2012).

The second key categorical distinction in the MPC lies between recklessness (R), which requires "awareness of the risk" that the circumstance exists ("suspicion"), and negligence (N), which supposes unawareness of the risk. To model attributions at the R-N boundary, we formulated scenarios that raised strong suspicion (R) and that raised only weak suspicion (N).

On the lower end of the culpability hierarchy, the distinction between negligence (N) and blamelessness (B) lies in whether the actor can fairly be blamed for not being aware of the risk under the circumstance. We tried to model that distinction by including or excluding facts or observations that ought to have raised suspicion (suggesting that the defendant might be blameworthy for failing to be alert to suspicious events). Across all of the mental states, the underlying questions are: How likely was it that the material circumstance existed, and was the actor blameworthy for taking the risk—i.e., for engaging in what turned out to be illegal conduct?

Experiment 1 demonstrated that subjects' attributions of mental state varied as expected depending on the strength of the evidence. Thus, the subjects appear to have grasped the basic conceptual distinctions being drawn by the MPC as we had modeled them.[34] The results from Experiment 1 indicate that with little to no training subjects can apply the MPC framework in a manner that is largely congruent with the basic assumptions of the MPC's mental state hierarchy. For example, in all four mental state categories, the modal response was what we characterized as the "best fit," and less than 10% of the subjects gave responses suggesting that they fundamentally misunderstood the task.[35] The findings in Experiment 1 gave us sufficient confidence to undertake the remaining experiments.

Experiment 1 also confirmed one of the key hypotheses of the study as well as a premise of the experimental design—that subjects would be rationally responsive to a request to attribute a "mental state" as defined in the MPC based on inferences drawn from "evidence" about the circumstances observed by the actor and descriptions of the actor's behavior. Specifically, we expected subjects to be willing to ascribe to the actor awareness of a fact, or suspicion about the existence of that

---

34. This observation is made only for purposes of evaluating our research method. It does not represent a normative statement about how we expect jurors to behave in criminal cases.

35. Given the amount of experimental "noise" we might expect for a study of this kind, these results are modestly encouraging about the ability of a typical person to interpret and apply the MPC hierarchy. While subjects sometimes made responses that indicated to us that they may have failed to understand critical legal distinctions presented to them (most notably, the 20% of individuals who labelled the blameless scenarios as constituting negligence), on the whole there were very few cases (less than 10%) where individuals provided responses indicating a failure to appreciate the basic framework of the MPC's mens rea provisions.

fact, even if the evidence did not show definitively that they knew it for sure. In other words, we expected our research subjects to carry out the assignment that real jurors are expected to carry out in any criminal trial where the defendant has contested whether or not she had the requisite mental state for a material circumstance. The findings in Experiment 1 confirmed these assumptions.[36]

B. From Attribution to Culpability

The most intriguing findings in this study emerged in Experiments 4–6. First, Experiment 4 (building on the intermediate steps taken in Experiments 2 and 3) shows that a strong suspicion that the material circumstance existed was regarded by a large majority of our subjects as a sufficient basis for culpability and punishment. Second, Experiments 5 and 6 show that when knowledge was specified as being necessary for conviction, strong suspicion that the material circumstance existed was regarded as sufficient "proof" of knowledge.

1. Intuitions About Culpability and Punishment

Beginning with Experiment 4, we examined how subjects conceived of the relationship between the MPC's mental state hierarchy and criminal culpability. The results from Experiment 4 can be characterized in two ways. First, most jury-eligible adults regard recklessness as a *necessary* basis for criminal liability. That is, they do not consider negligence to be a sufficient basis for criminalization in most of the scenarios that were used in this study.[37] In this respect, they share the intuition that led the MPC drafters to designate recklessness as the presumptive floor for criminal culpability, insisting on an explicit legislative justification to displace it with negligence.[38]

Second, Experiment 4 also indicates that for the typical jury-eligible participant in our experiments, recklessness regarding the existence of the circumstance is *sufficient* for holding the defendant

---

36. It is important to emphasize the potentially controversial nature of those findings, especially to readers who are puzzled, in an ontological sense, about whether it is *ever* possible to prove, based on contestable inference, that a person "knew" something (or even that they suspected something). Although this is not the place to take on such a profound question, it is the place to say that our subjects were clearly willing to make these attributions based on contestable evidence. This, of course, may be partly attributable to our design decision to exclude burdens of proof. One would expect noticeably different results if we had included an instruction that the mental state would have to be established "beyond a reasonable doubt." Future research can examine the extent to which this may be true.

37. Our subjects recognized the distinction between negligence and recklessness, and most apparently regarded negligence as an insufficient predicate for criminal liability.

38. MODEL PENAL CODE § 2.02(3) (AM. LAW INST. 1980).

criminally liable in these scenarios. In fact, there is no material difference in the proportion of subjects holding a defendant guilty when the evidence strongly suggests that he "knows" that the circumstance exists as compared to suspecting that it does. In other words, for the typical jury-eligible adult, there is a threshold for culpability and it exists not at knowledge, but at recklessness. This finding is especially intriguing in light of the fact that most criminal statutes (including drug possession, weapons possession, fraud, and identity theft) and even some civil statutes (like patent infringement) require knowledge as a necessary predicate for liability when the material circumstance differentiates lawful from unlawful conduct, as did all of our scenarios. To use the example above, though nearly all drug statutes require knowledge for conviction, for the average jury-eligible American, mere recklessness as to the presence of drugs in the bag is sufficient for conviction.

What can explain this stunning result? Two accounts come to mind. One line of explanation is that subjects are having trouble seeing a conceptual difference between reckless and knowing conduct and, because they cannot detect a difference between them, they are treated as morally equivalent. While this account is plausible, it is incompatible with the results of Experiment 1 and Experiment 3, which showed that subjects largely recognized the distinction between these mental states. The second possible explanation for the finding that recklessness is generally regarded as sufficient for criminal liability is that while subjects can differentiate recklessness and knowledge, they simply do not appreciate a *moral* distinction between them in relation to circumstance elements of criminal offenses. In the context of the scenarios used in this study, a substantial majority of our subjects regarded putative defendants who correctly suspected that the material circumstance existed as deserving of criminal punishment for "taking the risk" that they were committing a criminal act.

This finding naturally implicates the underlying normative question whether criminal statutes should require knowledge in some or all of these offenses, or whether recklessness should be regarded as a sufficient basis for punishment. The decision whether to require K or R is particularly interesting because in most litigated cases, unequivocal proof of knowledge is lacking and the key challenge is whether the circumstances were so suspicious that the actor must have "known" that that the material circumstance existed or, to use the MPC definition of knowledge,[39] that that there was a "high probability" that it existed. Given the MPC definitions of knowledge and recklessness,

---

39. *Id.* § 2.02(7).

attribution of "knowledge" to the actor rather than recklessness turns on whether there is a "high probability" that the circumstance existed or only a "substantial risk" that it did. In this context, one wonders not only about the determinacy of the distinction but also about the normative case for prescribing a mens rea of knowledge rather than recklessness.

Although the normative question reaches beyond our ambitions here, it is worth noting that the choice between recklessness and knowledge as a required mental state for material elements of offenses is a matter of genuine controversy. Even if recklessness is a morally sufficient basis for criminal liability, proof of knowledge might sometimes be required as a hedge against erroneous attributions of "conscious awareness of a risk" (i.e., suspicion). It is also noteworthy that many jurisdictions treat what is typically called "willful blindness" as a sufficient predicate for criminal punishment.[40] Under the doctrine of willful blindness, a reckless actor can be found guilty under a statute requiring knowledge insofar as the state can prove that the actor was aware of the risk that the material circumstance existed and took "deliberate actions to avoid learning of that fact."[41] The possibility of finding culpability without proof of "knowledge" can explain, in large part, the intuitive appeal of the doctrine of willful blindness. At the same time, however, some critics of the doctrine find proof of a conscious decision not to find out to be an ephemeral (and unnecessary) burden for the prosecution to bear in a case involving clearly reckless conduct,[42] while other critics find such proof *insufficient* to justify punishment when a statute requires knowledge.[43]

---

40. For a recent discussion of the case law, see Alexander F. Sarch, *Willful Ignorance, Culpability, and the Criminal Law*, 88 ST. JOHN'S L. REV. 1023 (2014).

41. Global-Tech Appliances, Inc. v. SEB S.A., 563 U.S. 754, 769 (2011). There is significant controversy, however, about what "deliberately avoiding knowledge" requires. *See, e.g.*, Deborah Hellman, *Willfully Blind for Good Reason*, 3 CRIM. L. & PHIL. 301 (2009); Douglas N. Husak & Craig A. Callender, *Wilful Ignorance, Knowledge, and the "Equal Culpability" Thesis: A Study of the Deeper Significance of the Principle of Legality*, 1994 WIS. L. REV. 29; David Luban, *Contrived Ignorance*, 87 GEO. L.J. 957 (1999); Alan C. Michaels, *Acceptance: The Missing Mental State*, 71 S. CAL. L. REV. 953 (1998); Sarch, *supra* note 40.

42. Robin Charlow, *Wilful Ignorance and Criminal Culpability*, 70 TEX. L. REV. 1351, 1358 (1992) (concluding that no current formulation of willful ignorance describes a mental state that is usually as reprehensible as knowledge); *see also* Alexander F. Sarch, *Beyond Willful Ignorance*, 88 U. COLO. L. REV. 97, 138–51 (2017) (arguing that willful ignorance should be expanded to include reckless ignorance).

43. Ira P. Robbins, *The Ostrich Instruction: Deliberate Ignorance as a Criminal Mens Rea*, 81 J. CRIM. L. & CRIMINOLOGY 191 (1990). However, Robbins recommends that legislatures sometimes replace a knowledge requirement with a requirement of recklessness. *Id.* at 231–34.

2. The Puzzling Impotency of the Knowledge Instruction

Experiment 5 demonstrates that the subjects were strongly inclined to regard recklessness as a sufficient predicate for liability *even when the subjects are instructed that "knowledge" is required under the statute.*[44]

What accounts for this eyebrow-raising observation? We can think of several likely possibilities. One is that jurors are engaging in a form of nullification: ignoring the statutory requirement for knowledge given their moral intuition that a reckless defendant is a guilty defendant. That is, jurors recognized that the defendant did not "know" that the material circumstance existed, but decided that conviction was warranted anyway. A second possibility is that the subjects were more inclined to attribute knowledge to the defendant in ambiguous factual scenarios when the statute prescribes knowledge as a necessary predicate for conviction. A third possibility is that the subjects were not attending to the statutory instruction despite the prominence we gave it in the task, reflecting a general tendency to ignore statutory language in favor of relying on bare intuitions regarding culpability. Finally, the problem could have been lack of clarity in our instructions. Experiment 6 sought to better parse which of these explanations, if any, might account for this result.

In Experiment 6, we found that nearly a third of those subjects who would have likely attributed recklessness to the actor in the absence of an instruction, attributed knowledge to the actor when they were instructed that the statute required knowledge as a basis for a finding of guilt. Thus, for many subjects, it would seem that the statutory language was serving as an anchor for their mental state attribution, and the attribution was perhaps being informed by their collective intuition that strong suspicion is a sufficient basis for a finding of culpability. In other words, the instruction, together with their moral intuitions about culpability, channeled their attributions of mental state. While giving the instruction led some subjects to attribute a different mental state than they might otherwise have chosen, it is unclear whether this reflected a conscious decision to ignore the instructions, thereby nullifying the statutory language. We are inclined to believe that the findings reflect a subconscious psychological tendency to infer knowledge when knowledge is required for conviction and the evidence revealed that the actor strongly suspected wrongdoing and made a blameworthy decision to undertake the forbidden conduct. But it is also possible that some participants may have consciously

---

44. For information on the instruction, see *supra* note 31 and accompanying text.

identified knowledge as being required for culpability under the statute and commensurately attributed that same mental state in order to provide concordance between their culpability determination, the statute, and their intuitions about what constitutes culpable conduct. The extent to which these two theories account for the observations presented here cannot be resolved using the present experimental design, and additional research is warranted.

C. Future Research

While numerous avenues of research are suggested by our results, three are of primary importance to our team.

1. Willful Blindness

While the MPC has promulgated the canonical mental states used in the present research, other mental states are actively used in a number of jurisdictions. One prominent example is willful blindness. Ongoing research by our group is exploring how the addition of willful blindness as a response option in tasks such as those described above shifts how participants categorize the offender's mental state. Early results indicate that most participants appear to see it as a natural explanation for offender conduct. If this finding holds up, its connection to the results presented above showing that recklessness is widely seen as the threshold for culpability will be explored in upcoming work.

2. The Effect of Hindsight Bias and Wrongful Conduct Bias

The scenarios used in this study test subjects' ability to distinguish someone else's knowledge of a fact from mere suspicion of its existence. How can a juror (or anyone else) ever know (much less "beyond a reasonable doubt") whether a criminal defendant knew or suspected something? These attributions have rarely been empirically studied in a legal context. Are they affected by known tendencies, biases or "rules of thumb"? Two such tendencies or biases are known to affect decisions of this nature and are worthy of further examination of the role they play in mental state attribution.

The first of these biases is hindsight bias. In our scenarios we always concluded the scenario by informing subjects that the circumstance in question did, in fact, exist. In this way our scenarios more closely parallel criminal cases typically presented to jurors.[45]

---

45. Inchoate offenses often involve situations where the circumstance did *not* exist. One example of such a case would be a charge for attempt when the victim was nonexistent. The most

However, by informing subjects that the circumstance actually existed it becomes impossible for subjects to objectively evaluate the probability it existed before that fact was known. This is known as hindsight bias and has been demonstrated to govern nearly all types of decisionmaking where it has been empirically examined. It has even been demonstrated to explain behavior of judges.[46] The extent to which juror knowledge of the outcome affects unbiased evaluation of the subjective and objective probability at the time of the offender's conduct is an interesting line of empirical and conceptual research.[47]

The second bias that may explain subjects' willingness to ascribe intentionality is known as the Knobe effect.[48] The Knobe effect explains the observed phenomenon that when a behavior has a negative externality, subjects are much more likely to ascribe intentionality to the actor with regard to that externality than when the behavior has a positive externality. In other words, by the very fact that the conduct we are speaking of here is wrongful and possibly harmful, we may be seeing that subjects are more willing to ascribe intentionality than they otherwise may be for neutral or positive conduct. Future studies can examine how subjects ascribe mental states with and without the information about whether the criminal circumstance actually exists, and perhaps with information that the circumstance does not actually exist. By doing so it may be possible to better parse the extent to which these biases affect the observed results.

3. Patterns and Correlates of Attributions

As expected, findings in all of our experiments reflected variations in attribution of mental states among the subjects. Unfortunately, our study was not designed to probe the patterns and correlates of these variations. However, we expect future research to explore the correlates of variations in attribution of specific mental

---

common example of such a case is a child sex abuse sting where the "victims" were in fact officers posing as young children. For a discussion of the mens rea considerations in cases such as these, see Donald S. Yamagami, *Prosecuting Cyber-Pedophiles: How Can Intent Be Shown in a Virtual World in Light of the Fantasy Defense?*, 41 SANTA CLARA L. REV. 547, 547–77 (2000).

46. *See, e.g.*, Chris Guthrie et al., *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1 (2007).

47. *See, e.g.*, Reid Hastie & W. Kip Viscusi, *What Juries Can't Do Well: The Jury's Performance as a Risk Manager*, 40 ARIZ. L. REV. 901 (1998).

48. Joshua Knobe, *Intentional Action and Side Effects in Ordinary Language*, 63 ANALYSIS 190, 190–94 (2003) (describing an experiment wherein the author asked individuals whether a CEO that made a business decision based purely on a financial basis intentionally harmed the environment when he had been told that environmental harm would be a byproduct of the business decision. The vast majority (82%) said yes. When provided with the same probe with the only change being that the outcome of the business decision helped the environment, fewer than one in four said that the businessman intended to help the environment).

states (e.g., K v. R and R v. N) and in the truly intriguing variations in subjects' intuitions about what level of awareness warrants punishment. Another important question is whether and to what extent jury-eligible subjects might vary in the willingness to attribute culpability depending on the relative strength of "suspicion" aroused by the evidence. For example, as noted above, one hypothesis is that people in general (or a subset, perhaps) may show hindsight bias in being more likely to attribute culpability (e.g., R) in ambiguous situations when the contested circumstance is known to have existed, as it was in all of our scenarios.

CONCLUSION

There are few more iconic moments in law than when jurors decide whether a defendant possessed a culpable state of mind—and if so which one—at the time of a bad act. But how do jurors actually decode past mental states of defendants? Despite the enormous consequences of these daily decisions for life and liberty, little, in fact, is known.

In six new exploratory studies, we empirically examined how jury-eligible adults went about deciding whether a hypothetical defendant—who had committed the actus reus of a criminal offense— also had the requisite "guilty mind" about a circumstance element in the definition of the offense. Our research focused on two specific and fundamental questions.

First, how do subjects go about drawing inferences about a person's mental state with respect to some legally relevant fact (such as the presence of contraband drugs inside a package) at the time the person committed an offense? Second, how do subjects evaluate the culpability of persons exhibiting different Model Penal Code mental states as to such facts? In answering both questions, we examined whether juror behavior and intuition were congruent with basic assumptions of the criminal law.

Briefly put, our experiments produced the following results.
➤ With little to no training, subjects appear to grasp the mens rea distinctions drawn by the MPC (at least as these distinctions pertain to circumstance elements of crimes). Moreover, subjects can apply those distinctions to fact patterns in a manner largely consistent with basic assumptions of the MPC hierarchy of culpable mental states. (Experiment 1).
➤ Subjects also appear to have a shared moral understanding of what level of culpability should be

criminally punishable—that is, what constitutes a "guilty mind." Strikingly, however, most subjects find a reckless mental state to be not only necessary for conviction, but also sufficient for it. (Experiments 2–4).

➤ Subjects are inclined to hold a defendant responsible when he is demonstrably aware of a *risk* that a legally relevant circumstance exists, even in the face of some uncertainty about whether he actually *knows* that it does. Moreover, and surprisingly, this inclination persists even when subjects are instructed that knowledge that the circumstance existed is legally required for conviction. (Experiment 5). This intriguing finding suggests that subjects are aligning their descriptive judgment about what mental state the actor possesses with their moral intuition that the actor is sufficiently culpable to be punished. (Experiment 6).

Our team is hesitant to draw sweeping conclusions from the results of these studies. And indeed even amongst ourselves we are in some disagreement about potential implications. Nonetheless, two general conclusions seem clear.

First, stakeholders in the criminal justice system can take some comfort that jury-eligible subjects appear to understand the MPC mens rea categories when it comes to circumstance elements, and that subjects are largely able to apply those categories in the way the drafters of the MPC, and legislatures adopting the MPC, expected.

Second, this comfort comes with a caveat: subjects not only find (unsurprisingly) that a defendant's awareness of a significant risk that a legally relevant circumstance exists is sufficient for criminal culpability, they also conclude that the same level of awareness is sufficient in the face of a statute that permits conviction only if the defendant knows that it exists. Put another way, when faced with a tension between their moral intuition that what the law calls recklessness is enough for conviction, on one hand, and the existence of statutory language that requires a defendant's knowledge, on the other, subjects tend simply to *find* that the actor possessed knowledge, even when the contexts arguably demonstrate that the actor was only reckless.

More work will be needed, clearly, to fully mine the implications of these findings. In the meantime, however, our research could have important consequences for legislatures, appellate courts, and trial judges who are trying to decide whether a defendant needs to *know* a

fact to be criminally responsible, or merely needs to be reckless or negligent.

Criminal law scholars have long debated the mental state and culpability criteria that should be necessary and sufficient for criminal liability. But they have paid too little attention to the question of how a jury actually draws the culpability distinctions that are theoretically relevant if the criminal law is to perform its retributive, deterrent, incapacitating, and expressive functions. Our novel empirical studies squarely address that question and offer some highly intriguing, albeit preliminary, answers.

APPENDIX A: THE SUBJECTS

Individuals were recruited to participate in our experiments between November 2015 and October 2016. All recruitment and experimental procedures were approved by the Vanderbilt Institutional Review Board.[49] All recruitment was done online using Amazon's Mechanical Turk service ("AMT"). AMT is a marketplace where individuals from across the globe can perform various tasks for payment from various providers. Research using these web-based recruiting techniques has been widely validated and also provides samples of the population that are substantially more representative than samples of convenience typically used in such studies.[50] Nonetheless, there is always the possibility, when subjects are participating remotely, that they may not be fully attending to the task at hand. In order to account for this, we excluded subjects who took an abnormal amount of time to respond.[51] We also included an attentional check that screened out subjects who did not read the material closely[52] and we only used subjects who had an established history of satisfactorily completing tasks on AMT.[53]

Subjects agreeing to participate through AMT were then directed to the experiment, which was hosted by Qualtrics. Qualtrics is a web-based platform for hosting surveys and experiments. It is regularly used by scholars in many fields and was also used by this group of authors for our previous two studies described in Section III.A,

49. The Vanderbilt University Institutional Review Board ensures that approved studies are in compliance with both federal guidelines and widely held norms for human experimentation.

50. *See, e.g.*, Tara S. Behrend et al., *The Viability of Crowdsourcing for Survey Research*, 43 BEHAV. RES. METHODS 800 (2011); Adam J. Berinsky et al., *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, 20 POL. ANALYSIS 351 (2012); Michael Buhrmester et al., *Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?*, 6 PERSP. ON PSYCHOL. SCI. 3 (2011); Joseph K. Goodman et al., *Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples*, 26 J. BEHAV. DECISION MAKING 213 (2013); Jon Sprouse, *A Validation of Amazon Mechanical Turk for the Collection of Acceptability Judgments in Linguistic Theory*, 43 BEHAV. RES. METHODS 155 (2011).

51. Subjects whose timing was two standard deviations faster or slower than the average participant were excluded. This is a customary screening technique to detect noncompliance with task instructions.

52. In addition to the fact patterns contained in Appendix B, subjects were presented with a fact pattern that looked like a normal fact pattern but had a sentence in the middle of the paragraph that read, "This is an attention check. Please ignore the rest of this text and select a response of [specific response varied depending on the experiment] to confirm that you are reading these scenarios carefully." All subjects selecting a response other than the directed response were excluded from the analysis.

53. Note that in this regard the steps that we took to ensure the complete attention of our subjects may be more rigorous than those methods used by courts.

above.[54] At the conclusion of their participation subjects were debriefed and paid.

Before participating in our studies, Amazon confirmed that all potential participants were U.S. citizens or residents over the age of eighteen by means of a U.S. based bank account as well as their IP address.[55] All experiments ended with subjects providing some demographic information, which allowed us to confirm that the sample was generally representative of the U.S. jury-eligible population. We provide the summary statistics of this sample as compared to the U.S. population in Table A1. Table A2 then lists the total number of subjects per experiment.

---

54. Studies relying on Qualtrics experiments include Jonathan S. Abramowitz et al., *Obsessive–Compulsive Symptoms: The Contribution of Obsessional Beliefs and Experiential Avoidance*, 23 J. ANXIETY DISORDERS 160, 162 (2009); Yany Grégoire et al., *When Customer Love Turns into Lasting Hate: The Effects of Relationship Strength and Time on Customer Revenge and Avoidance*, J. MARKETING, Nov. 2009, at 18, 21; and Paul H. Robinson et al., *The Disutility of Injustice*, 85 N.Y.U. L. REV. 1940, 2000, 2004 (2010).

55. Subject totals for each experiment were based on power analyses that indicated the totals that would be sufficient for us to either confirm or reject the hypothesis being tested. Power analyses allow for experimenters to determine the probability of a false negative (that is, concluding no difference exists when one actually does) and reduce this probability to widely accepted levels.

TABLE A1: DEMOGRAPHICS OF EXPERIMENTAL
SUBJECTS

Education	Subjects	U.S. Census
Less than HS	<1%	18%
High school / GED	10%	30%
Some college	26%	20%
Assoc. degree	13%	7%
Bachelor's	38%	17%
Graduate Degree	13%	10%

Income	Subjects	U.S. Census
< $20k	25%	$1 to $25k: 22%
$20k - $40k	30%	$25k to $35k: 19%
$40k - $60k	21%	$35k to $50k: 21%
$60k - $80k	9%	$50k to $65k: 14%
$80k - $100k	11%	$65k to $75k: 6%
> $100k	5%	$75k to $100k: 8%

Gender	Subjects	U.S. Census
Male	53%	49%
Female	46%	51%

Age Groups	Subjects	U.S. Census
18-24	12%	13%
25-34	46%	18%
35-44	22%	18%
45-54	12%	19%
55-64	6%	16%
65+	2%	18%

Race	Subjects	U.S. Census
White	77%	74%
Non-White	23%	26%

TABLE A2: NUMBER OF SUBJECTS BY EXPERIMENT

Experiment	*Number of Subjects*
1. Attributing Mental States Using MPC Definitions	118
2. Subjects' Unguided Moral Intuitions About What Mental State Is Sufficient for Criminal Liability	95
3. Use of Signals to Align Subjects' Responses with the Best Fit Mental State	119
4. Subjects' Unguided Intuitions About Culpability Sufficient for Criminal Liability: Revisited	99
5. The Effect of Instructions	103
6. The Effect of Instructions: Revisited	94
TOTAL SUBJECTS, ACROSS ALL EXPERIMENTS	*628*

APPENDIX B: FULL TEXT OF SCENARIOS

Appendix B details the full set of sixty-three scenarios used in the experiments. Due to its length, it is provided online at this location: http://www.lawneuro.org/Appendix_B.pdf   [https://perma.cc/8SDD-SBGP].

***