SORTING GUILTY MINDS

FRANCIS X. SHEN,[α] MORRIS B. HOFFMAN,[β] OWEN D. JONES,[χ] JOSHUA D. GREENE,[δ] & RENÉ MAROIS[ε]

Because punishable guilt requires that bad thoughts accompany bad acts, the Model Penal Code (MPC) typically requires that jurors infer the mental state of a criminal defendant at the time the crime was committed. Specifically, jurors must sort the defendant's mental state into one of four specific categories—purposeful, knowing, reckless, or negligent—which will in turn define both the nature of the crime and the degree of the punishment. The MPC therefore assumes that ordinary people naturally sort mental states into these four categories with a high degree of accuracy, or at least that they can reliably do so when properly instructed. It also assumes that ordinary people will order these categories of mental state, by increasing amount of punishment, in the same severity hierarchy that the MPC prescribes.

The MPC, now turning fifty years old, has previously escaped the scrutiny of comprehensive empirical research on these assumptions underlying its culpability architecture. Our new empirical studies, reported here, find that most of the mens rea assumptions embedded in the MPC are reasonably accurate as a behavioral matter. Even without the aid of the MPC definitions, subjects were able to distinguish regularly and accurately among purposeful, negligent, and blameless conduct. However, our subjects failed to distinguish reliably between knowing and reckless

---

[α] Visiting Assistant Professor, Tulane University Law School and The Murphy Institute.

[β] District Judge, Second Judicial District (Denver), State of Colorado; Adjunct Professor of Law, University of Colorado; Member, John D. and Catherine T. MacArthur Foundation Research Network on Law and Neuroscience; Research Fellow, Gruter Institute for Law and Behavioral Research.

[χ] New York Alumni Chancellor's Chair in Law & Professor of Biology, Vanderbilt University; Director, John D. and Catherine T. MacArthur Foundation Research Network on Law and Neuroscience.

[δ] Associate Professor of Psychology and Director of the Moral Cognition Laboratory, Harvard University.

[ε] Associate Professor of Psychology and Director of the Human Information Processing Laboratory, Vanderbilt University. We received helpful comments from Al Alschuler, Sara Beale, Stephanos Bibas, Ted Blumoff, Richard Bonnie, Josh Dressler, Nita Farahany, Jeff Fagan, Pat Furman, Ken Gallant, David Gates, Dena Gromet, Dan Kahan, Rob Mikos, Thomas Nadelhoffer, Bill Pizzi, Jeff Rachlinski, Fred Schauer, Kenneth Simons, and Chris Slobogin, as well as from the organizers of and participants in the Adjudicating Guilty Minds Symposium at Duke Law School, the Guilty Minds: Neuroscience & Criminal Law Symposium of the Denver University Law Review, and conferences of the MacArthur Foundation Law and Neuroscience Project. Jonathan Dial, Katherine Jan, Katherine Kuhn, Tim Mitchell, Cameron Munier, and Sarah Pazar provided valuable research assistance. Preparation of this Article was supported by the John D. and Catherine T. MacArthur Foundation (Grant # 07-89249-000 HCD), the Regents of the University of California, the Center for Integrative and Cognitive Neuroscience (CICN), and Vanderbilt University.

conduct. This failure can have significant sentencing consequences for certain crimes, especially homicide.

INTRODUCTION

In its dark and quiet core, the administration of criminal justice in America depends—far more than we like to admit—on amateur mind readers. This is because the thought processes accompanying an act dramatically affect our assessments of blameworthiness and our subsequent decisions to punish. We care, for example, whether a shooter intended to shoot and injure the person he killed. We therefore ask jurors to infer the mental state of a defendant they do not know as he acted in a way they did not see.

Specifically, jurors must sort the defendant's mental state into one of four defined categories. This is because the vast majority of states either have adopted or have been heavily influenced by the

Model Penal Code (MPC),[1] which since 1962 has divided the universe of potential culpable mental states into: (1) purposeful; (2) knowing; (3) reckless; and (4) negligent.[2]

This MPC taxonomy reflects several assumptions. It assumes that average people naturally sort real-world mental states into these four categories with reasonable reliability—or at least that they can when so instructed. Of course, not every criminalized act generates four different levels of defined crime. Quite often, a particular crime is defined exclusively as a given act committed with a specific level of mental state. Acts committed with more culpable mental states are punished no more severely, and acts committed with less culpable mental states are not crimes at all. The MPC has formalized this idea by providing, in § 2.02(5), that if a crime requires a certain mental state (say, recklessness) then a person who commits the act with a more culpable mental state (knowledge or purpose) is still guilty of the crime.[3] On the other hand, there are some serious crimes, such as homicide, which are typically defined by differing degrees of culpability. Those differing degrees make a purposeful act more serious than a knowing act, a knowing act more serious than a reckless act, and so on. For such crimes, the MPC further assumes that, holding the act constant, the average person would punish these four categories in the manner corresponding to the MPC hierarchy—that is, punishing *purposeful* conduct the most severely and *negligent* conduct the least severely.

We are now in year fifty of the MPC dynasty. Given the dramatic consequences that assigning different mental states can generate in the criminal justice system, you might think that, fifty years in, the underlying culpability assumptions of the MPC had been rigorously and empirically tested. But you would be wrong. With only a few methodologically unsatisfying exceptions, there is no empirical literature on the validity of the MPC culpability assumptions. The dearth of empirical studies is all the more striking given that the MPC is both the principal text for teaching students criminal law and the most widespread regime that criminal defendants encounter as they are tried, convicted, and sentenced.

With a grant from the MacArthur Foundation, we set out to investigate these critical, yet so far insufficiently tested, MPC assumptions. Assembling an interdisciplinary team of legal scholars and scientists, we designed and conducted the first comprehensive series

---

[1] 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.

[2] MODEL PENAL CODE § 2.02 (1962).

[3] *Id.* § 2.02(5).

of experiments, on which we report here, to address the validity of the MPC culpability assumptions. The bottom line emerging from our analysis is that in almost all of our experimental conditions, subjects behaved as the MPC assumes they would, with or without the assistance of jury instructions. But one very important exception emerged at the boundary between knowing and reckless conduct. In assigning punishment, subjects were less able to differentiate between knowing and reckless conduct, even with the benefit of jury instructions. While our results largely validate the MPC approach to culpability, the difficulty at the knowing/reckless boundary may suggest a need for reform. Such reform might include improving jury instructions, redefining the knowledge and reckless categories, or abandoning the distinction between those two categories either entirely or only in homicide cases, where the MPC distinction can have dramatic consequences.

Part I of this Article provides, for context, a brief history of culpability theories and the Model Penal Code. Part II describes the scant existing empirical literature on juror assessments of MPC mental states. Part III details the design and results of our experiments. Part IV discusses the implications of our results, including some narrow areas for possible reform. Appendix A provides technical details of the experiments, and Appendix B provides the full text of all 150 scenarios used in the experiments.

I
Culpability and the Model Penal Code

Accidents happen, and it seems to be a human universal that we generally do not punish truly accidental acts, but only culpable ones. The idea of culpability has existed as long as humans have punished each other. Primitive societies, both ancient and extant, universally seem to recognize a moral difference between accidents and non-accidents.[4] Every ancient civilization that has left a record on the

[4] *See, e.g.*, E. Adamson Hoebel, The Law of Primitive Man 235–36 (1954) (discussing the Ashanti law of homicide); Paul Radin, The World of Primitive Man 248–51 (1960) (discussing the Bantu conception of culpability). While there are many examples throughout history of strict liability crimes, these seem to be the exception rather than the rule. Paul H. Robinson, *A Brief History of Distinctions in Criminal Culpability*, 31 Hastings L.J. 815, 823–25 (1980). For example, property owners were often held strictly liable for damage caused by their property, including the acts of their slaves. 2 Frederick Pollack & Fredric William Maitland, The History of English Law 470–73 (2d ed. 1968). A similar idea is expressed in the modern law of products liability. Sometimes the property itself was blamed, as with the ancient Norse and early English doctrine of deodand, under which property that injured others was destroyed. Some scholars have argued that deodand explains, in part, our current rule that a legal fiction like a corporation, which

issue—including the Babylonians, Jews, Egyptians, Greeks, and Romans—has recognized that blameworthy wrongs must usually have some component related to the wrongdoer's state of mind in order to distinguish them from pure accidents.[5] The English precept from which we get our phrase "mens rea" ("guilty mind") was "*actus non facit reum nisi mens sit rea*,"[6] which means "an act is not guilty unless the mind is guilty."[7]

But can we slice intentionality more finely than accident versus non-accident?[8] Do we believe that there are morally relevant distinctions within the general category of "intentionality"? Similarly, are some kinds of accidents more blameworthy than others? Such questions have long vexed the law.

At the accident end of this spectrum, it seems we always have made a distinction between careless and blameless accidents. The legal roots of this distinction appear to be as old and universal as the accident/non-accident distinction itself, going deep into both the Roman and Anglo-Saxon-German branches of the common law.[9] Non-European societies also appear to have recognized this difference. Bantu tribesmen in South Africa recognized it long before contact with Europeans, as did the Jalé of New Guinea, though only through informal procedures, to name only two pre-industrial societies.[10] This distinction recognizes that while accidents happen, some accidents happen because people are not as careful as they should be.

---

has no mind and therefore cannot have any state of mind, can nonetheless be held criminally liable. Albert W. Alschuler, *Two Ways To Think About the Punishment of Corporations*, 46 AM. CRIM. L. REV. 1359, 1392 (2009).

[5] Max Radin, *Intent, Criminal*, *in* 8 ENCYCLOPAEDIA OF THE SOCIAL SCIENCES 126, 126–27 (Edwin R.A. Seligman & Alvin Johnson eds., 1932) (indicating that intention generally was required under Pentateuchal, Greek, and Roman law). For example, Hammurabi's Code provided: "If during a quarrel one man strike another and wound him, then he shall swear, 'I did not injure him wittingly,' and pay the physicians." THE CODE OF HAMMURABI 44 (L.W. King trans., 2007).

[6] Francis Bowes Sayre, *Mens Rea*, 45 HARV. L. REV. 974, 988 (1932). This famous phrase dates at least from the time of Henry I in the early 1100s, but was likely based on the writings of St. Augustine. *Id.* at 983 & n.30.

[7] *See* 4 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 21 (1769) ("[A]n unwarrantable act without a vitious will is no crime at all.").

[8] We use the word "intentionality" to mean culpable mental states, and we use intentionality interchangeably with the word "culpability." For crimes that require a guilty mind, culpability is synonymous with intentionality because—holding constant the criminal act—the actor's level of culpability is based on his level of intentionality.

[9] *See* Robinson, *supra* note 4, at 825–30 (discussing the willful/accidental distinction in the history of the common law).

[10] 1 RALPH PIDDINGTON, AN INTRODUCTION TO SOCIAL ANTHROPOLOGY 345, 349 (1st ed. 1950) (Bantu); HORIZONS OF ANTHROPOLOGY 316 (Sol Tax & Leslie G. Freeman eds., 2d ed. 1977) (Jalé); Robinson, *supra* note 4, at 850 (same).

This first cut is what we call, in modern parlance, the difference between negligent and non-negligent acts. Although that difference now animates the law of torts, it has been part of the notion of blameworthiness long before the existence of the modern distinction between crime and tort.[11] While there are still a handful of crimes based on negligence, they are the exception rather than the rule.[12]

For the lion's share of crimes that require some level of culpability beyond negligence, history's next cut was to recognize a difference between mere negligence and something that various legal systems called "recklessness," "gross negligence," "willful blindness," or other labels that connote a level of culpability higher than mere negligence but lower than the infliction of desire-based harm.[13] Such behavior is arguably different from mere negligence in that the purely negligent actor has no consciousness of the risk to which he is exposing others; indeed, the essence of mere negligence is the failure to appreciate that risk. But, the argument continues, when an actor has some appreciation of the risk of harm, yet takes that risk anyway to achieve some other desired result, he is behaving in a manner qualitatively different from, and more deserving of punishment, than if he were just inattentive.

The roots of this distinction clearly predate the common law and are seen in several ancient societies.[14] The gist of this lower level of

---

[11] With a few noteworthy exceptions like Hammurabi's Code and the laws of Moses, the criminal law as we think of it today—comprehensively governing virtually all wrongs committed by one individual against another—is a relatively recent invention. In most ancient societies, and with the exception of certain crimes against the state like regicide and treason, the state simply did not get involved with the behaviors of individuals, who were left to resort to private revenge. *See* James Lindgren, *Why the Ancients May Not Have Needed a System of Criminal Law*, 76 B.U. L. REV. 29, 33–36 (1996) (indicating that in many ancient societies, the role of the state was at most to provide a forum for individuals to resolve private disputes, rather than to formulate and enforce substantive criminal laws).

[12] For example, the MPC criminalizes negligent homicide. MODEL PENAL CODE § 210.4 (1962). It seems that when the harm is great, we are more willing to criminalize unintentional but negligent acts. Much of the pre-MPC controversy about culpability centered on the question of when merely negligent acts should be criminalized. *See generally* ROY MORELAND, A RATIONALE OF CRIMINAL NEGLIGENCE (1944) (surveying law of negligent homicide).

[13] Robinson, *supra* note 4, at 837–46.

[14] One of the earliest Anglo-Saxon descriptions of this kind of "negligence-plus," contained in the *Laws of King Alfred*, was lifted almost verbatim from Mosaic Law as described in the Book of Exodus:

> If an ox gore a man or a woman, so that they die, let it be stoned, and let not its flesh be eaten. The lord shall not be liable, if the ox were wont to push with its horns for two or three days before, and the lord knew it not; but if he knew it, and he would not shut it in, and it then shall have slain a man or a woman, let it be stoned; and let the lord be slain . . . .

ANCIENT LAWS AND INSTITUTES OF ENGLAND 22 (B. Thorpe ed., 1840).

intentionality, and higher level of negligence, was that it is wrong for a person to harm another by taking an inordinate risk—sufficiently wrong to be criminal. The critical idea here is that, although we may want to punish these sorts of unintended acts, we punish them less harshly than intended harms. This recognizes the distinction, found in moral philosophy since Aquinas, that intended harms are more culpable than harmful side effects.[15]

The most recent major fault line in the law of intentionality seems to be a purely American invention, distinguishing between desire-based intent and a new category of "recklessness-plus."[16] This new distinction, grounded in the degree of risk the actor is consciously undertaking, attempts to capture the situation where a particular harm is not desired but is nevertheless virtually certain to occur if the actor acts. Such a state of mind (now called "knowing" by the MPC) seems less blameworthy than pure desire-based harms but more blameworthy than merely taking a lesser risk ("reckless" in MPC language).[17] It is one thing (reckless) for me to shoot over a victim's head to kill a bird, killing the victim instead, and perhaps another (knowing) to shoot through the victim to kill the bird. In both cases, the wrong is the conscious disregard of a known risk, but in the former case the risk is something shy of 100% while in the latter it is effec-

[15] John Finnis, *Object and Intention in Moral Judgments According to Aquinas*, 55 THOMIST 1, 1–3 (1991). The experimental philosophy literature on intentionality reminds us that side effects are complicated. Even if a harm is a side effect (as opposed to a direct effect) of a given action, if the actor knew that the side effect would occur, but acted anyway, we will tend to judge the actor as if he intended to cause the side effect. *See* Joshua Knobe, *Intentional Action and Side Effects in Ordinary Language*, 63 ANALYSIS 190, 190–93 (2003) (discussing an experiment in which subjects treated unintended, but foreseen, side effects as though they were intentional). For discussion of the doctrine of the double effect, in which intended harm as a means to a certain end is seen as morally worse than equivalent harm foreseen as a side effct of an end, see generally Fiery Cushman, Liane Young & Marc Hauser, *The Role of Conscious Reasoning and Intuition in Moral Judgment*, 17 PSYCHOL. SCI. 1082 (2006).

[16] Robinson, *supra* note 4, at 846–49.

[17] "Knowingly" as a culpability category is new only as it relates to so-called "results" elements of offenses; this category has long been part of the criminal law as it relates to so-called "circumstances" elements, or, as the MPC calls them, "attendant circumstances." MODEL PENAL CODE § 1.13(9). In addition to, or even instead of, a harm element, some crimes contain elements that require a particular mental state as to an existing or historical fact. For example, the federal crime of sending or receiving child pornography in interstate commerce, see *infra* note 26, requires the defendant to "know" both that the material is pornographic and that the persons depicted in it are children. "Knowing" a circumstance element is substantially more straightforward than "knowing" about a risk of future harm, which probably explains why the law has long recognized "knowing" as a circumstances state of mind, but only recently recognized it as a results state of mind. Our experiments looked only at mental states as they relate to results elements, and in the balance of this Article, when we discuss a particular mental state, we refer to that mental state as applied only to results elements.

tively 100%.[18] There appear to be no express articulations of this new recklessness-plus in any legal systems prior to its first suggestion in the 1830s in an American treatise on federal Indian law.[19]

Other smaller fissures in the culpability continuum have suggested themselves over the centuries. Homicide, no doubt because it was considered in most systems to be the most serious of all crimes, seems to have been a particularly prolific generator of additional state-of-mind categories. Doctrines with names like "heat of passion" and "provocation," though technically applicable to many criminal offenses, were almost exclusively born from and applied in homicide cases, with the result that these doctrines blurred even further the grades of homicide based on different states of mind.[20] Murder even has its own category of super-intentionality, at least in the United States. In virtually every state, first degree murder, penalized by the most serious of punishments, whether life in prison or the death penalty, requires not just an intentional killing, but a killing carried out after deliberation or with premeditation.[21]

---

[18] The distinction between purposeful and knowing will not matter in most criminal codes because, like the common law, most codes recognize a special kind of reckless homicide that is so gross, and so insensitive to the risk of killing, that the act will be treated as if it were comitted purposefully. Most jurisdictions have followed the MPC lead, at § 210.2(1)(b), in calling this specially heightened form of recklessness "extreme indifference," though the common law used more flowery descriptions, such as "evincing a depraved heart, devoid of social duty, and fatally bent on mischief." WAYNE R. LAFAVE, CRIMINAL LAW 739–40 (4th ed. 2003). By whatever name, the knowing bird-shooter is guilty of first degree murder in most jurisdictions, even though his purpose was not to kill any person.

[19] Robinson, *supra* note 4, at 846.

[20] Joshua Dressler, *Rethinking Heat of Passion: A Defense in Search of a Rationale*, 73 J. CRIM. L. & CRIMINOLOGY 421, 447–48 (1982); Stephen J. Morse, *Diminished Rationality, Diminished Responsibility*, 1 OHIO ST. J. CRIM. L. 289, 296 (2003).

[21] For examples of state first degree murder statutes containing premeditation or deliberation as an element, see ARK. CODE ANN. § 5-10-101 (2011); CAL. PENAL CODE §§ 187–188 (Deering 2011); COLO. REV. STAT. § 18-3-102 (2010); D.C. CODE § 22-2101 (LexisNexis 2011); GA. CODE ANN. § 16-5-1 (2011); IDAHO CODE ANN. §§ 18-4001 to -4002 (2011); IOWA CODE § 707.2 (2010); MD. CODE ANN., CRIM. LAW § 2-201 (LexisNexis 2011); MASS. ANN. LAWS ch. 265, § 1 (LexisNexis 2011); MICH. COMP. LAWS SERV. § 750.316 (LexisNexis 2011); MISS. CODE ANN. § 97-3-19 (2010); MO. REV. STAT. § 565.020 (2011); MONT. CODE ANN. § 45-5-102 (2010); NEB. REV. STAT. ANN. § 28-303 (LexisNexis 2010); NEV. REV. STAT. ANN. § 200.030 (LexisNexis 2011); N.H. REV. STAT. ANN. § 630:1-a (LexisNexis 2011); N.M. STAT. ANN. § 30-2-1 (LexisNexis 2010); N.C. GEN. STAT. § 14-17 (2011); OKLA. STAT. tit. 21, § 701.7 (2011); 18 PA. CONS. STAT. § 2502 (2010); R.I. GEN. LAWS § 11-23-1 (2011); VT. STAT. ANN. tit. 13, § 2301 (2011); VA. CODE ANN. § 18.2-31 (2011); W. VA. CODE ANN. § 61-2-1 (LexisNexis 2010); WYO. STAT. ANN. § 6-2-101 (2010). Interestingly, there is no second degree murder in England—the alternatives are murder for an intentional killing with or without deliberation (punished by a life sentence) and manslaughter for everything else (punished by lesser penalties). Parliament's Law Commission is considering distinguishing between first and second degree homicide, as is done in American criminal law, with deliberation as the distinguishing factor. Tom

Of course, the full history of the law of culpability is considerably more confusing, and less linear, than the brief rendition in the preceding paragraphs might suggest. When governments in general, and the common law in particular, began to grapple with the problem of private wrongs, they did so haltingly and inconsistently. At some times in some systems, "intentional" still meant anything that was not an accident.[22] But at other times, various systems tried to tease apart intentionality into the different varieties summarized above.[23] Even when they did, different categories of intentionality appeared at different times, and were described differently by different legal systems, and even by different courts in the same system. Definitions overlapped and conflicted. Culpability mattered for some crimes and not for others.[24]

If this cacophony were not bad enough, the deeply complicated question of whether the culpability inquiries should be subjective or objective only multiplied the variations and confusion. When we ask whether a generic defendant John was reckless, are we asking a question about John's subjective state of mind, or an objective question about how most of us would have acted in John's place? The common law answered this question in wildly inconsistent ways. It generally pretended to treat the question subjectively, as if asking what was inside a criminal's mind at the time of the crime were a factual inquiry not unlike what was inside a safe deposit box.[25] But in practice its subjective-sounding inquiries always had irreducibly objective strands, because of course judges and jurors cannot get inside the criminal's mind to see what he intended at the time of his crime. Thus, when we ask ourselves what was in John's mind, we end up asking what our own mental state would be if faced with John's situation.[26]

Whitehead & Laura Roberts, *Murderers 'To Escape' Automatic Life Sentences*, THE TELEGRAPH (July 12, 2010), http://www.telegraph.co.uk/news/uknews/law-and-order/7886251/Murderers-to-escape-automatic-life-sentences.html.

22 Robinson, *supra* note 4, at 825–30.

23 *Id.* at 833–49.

24 *See* Sayre, *supra* note 6, at 1016 (concluding that mens rea has had "no fixed continuing meaning" and "has varied with the changing underlying conceptions and objectives of criminal justice").

25 *See* John Shepard Wiley, Jr., *Not Guilty by Reason of Blamelessness: Culpability in Federal Criminal Interpretation*, 85 VA. L. REV. 1021, 1064 (1999) ("Persons trained only in the common-law tradition (including Supreme Court Justices until very recently) often thought the question of culpability meant requiring actual, subjective knowledge or requiring nothing.").

26 Even when the law has settled on a given state of mind for a given crime and tried to solve the subjective/objective problem, it has remained unclear as to whether that state of mind must apply to all the elements of the defined crime. Imagine that a jurisdiction has defined the crime of trafficking in child pornography as "knowingly transporting, receiving or distributing in commerce any visual depiction of a minor engaging in sexually explicit

These are extraordinarily difficult intellectual issues in their own right, and they are only exacerbated when political bodies are called upon to address them. Criminal code drafting in the United States was a major part of the legislative agenda of states for the first half of the 1800s, but then American legislatures essentially fell silent about general criminal law principles for the next 100 years.[27] By 1950, this abject neglect left state criminal codes in what the Supreme Court famously described as "disparity and confusion [over the] definitions of the requisite but elusive mental element" of crimes.[28] Commentators were less restrained, one describing state criminal codes as "archaic, inconsistent, unfair, and unprincipled."[29] Congress did no better. It began in the early 1900s to federalize many aspects of the existing criminal law and to define entirely new federal crimes, and it has never slowed down. But Congress has never attempted to answer these beguiling culpability questions. To this day, the federal criminal code contains no general culpability definitions.[30]

It was this horribly unsettled state of the law of culpability that confronted the American Law Institute (ALI), a collection of widely respected lawyers, judges, and academics, when it began to consider

---

conduct." Now imagine that John is arrested as he transports a child pornography video. John admits he "knowingly" transported the video, and admits that he knew it was pornographic, but claims he did not "know" the person in the video was a minor. That is, John argues that the word "knowingly" modifies each and every one of the elements of the act, and that because he did not know the subject was a minor, he cannot be convicted of this offense. These were the facts that faced the Supreme Court in *United States v. X-Citement Video, Inc.*, 513 U.S. 64 (1994). The Court held that the manner in which Congress chose to define this particular crime did in fact mean that the mental state requirement applied to each element, including the age of the subject, and therefore reversed the conviction. *Id.* at 78. This problem is really a specific example of the more general question of statutory construction. Both state and federal courts have generally followed the principle that if a statute has only one state of mind listed at the beginning of the definition, that state of mind applies to all the following elemental acts. JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 136–37 (5th ed. 2009). But legislative intent is not always clear, and thus this issue continues to be an interpretive crapshoot, often requiring courts to guess at the core nature of the crime the legislative body was trying to reach.

[27] DRESSLER, *supra* note 26, at 30.

[28] Morissette v. United States, 342 U.S. 246, 252 (1952).

[29] Sanford H. Kadish, *Fifty Years of Criminal Law: An Opinionated Review*, 87 CALIF. L. REV. 943, 947 (1999); *see also* Herbert Wechsler, *The Challenge of a Model Penal Code*, 65 HARV. L. REV. 1097, 1100–01 (1952) (discussing legislative and judicial "neglect" of substantive criminal law over the last twenty years).

[30] Instead, each defined federal crime either contains its own culpability requirement (often in non-MPC language, such as the "malice aforethought" or "willful, deliberate, malicious, and premeditated" killing required for first degree murder under 18 U.S.C. § 1111(a) (2006)) or contains no culpability requirement at all. When there is no express mental state requirement, courts must infer one. The absence of any general culpability provisions under federal statutory law has forced the Supreme Court to develop its own culpability jurisprudence, with decidedly mixed results. *See* Wiley, *supra* note 25, at 1023 (arguing that the Supreme Court has equated criminal culpability with moral culpability).

criminal law reform in the 1950s. Led by Herbert Wechsler of Columbia Law School, the ALI undertook to do what no legal system had ever expressly tried to do: orchestrate the noise of culpability into a reasonably uniform and workable system. After thirteen tentative drafts and accompanying commentaries, the ALI published its first Model Penal Code in 1962. It addressed three broad areas sorely in need of attention: sentencing, the definition and classification of specific crimes, and, most important for our purposes, general principles of criminal responsibility.

The MPC settled on four categories of criminal responsibility, which it called (1) purposeful (and which some jurisdictions call intentional); (2) knowing; (3) reckless; and (4) negligent. It defined them this way:

> A person acts purposely [with respect to a result] if it is his conscious object . . . to cause such a result.
>
> A person acts knowingly [with respect to a result] if . . . he is aware that it is practically certain that his conduct will cause such a result.
>
> A person acts recklessly [with respect to a result] when he consciously disregards a substantial and unjustifiable risk that [his conduct will cause the result].
>
> A person acts negligently [with respect to a result] when he should be aware of a substantial and unjustifiable risk that [his conduct will cause the result].[31]

The ALI retained the oldest culpability distinctions—between purposeful, negligent, and blameless, and between negligent and reckless—and also retained the newest distinction between reckless and knowing. It declined to slice culpability any further as a general matter.[32] It also set out, in its definitions of specific crimes, a general architecture that requires that every crime consist of both an act and one of the four levels of culpability.[33]

---

[31] MODEL PENAL CODE § 2.02 (1962). For the definitions of purposeful and knowing, we have omitted the alternate definitions for when those mental states are applied to circumstance elements, as opposed to results. *See supra* note 17 (discussing mental states applied to circumstance elements).

[32] Although the MPC retained some of the more finely grained common law subspecies of culpability, it did so by incorporating these culpability levels into the definitions of specific crimes, rather than as stand-alone levels of culpability. So, for example, it retained the common law concept of a homicide committed in the "heat of passion" (though it uses the phrase "under the influence of extreme mental or emotional disturbance") as part of the definition of manslaughter. *Id.* § 210.3(1)(b).

[33] *See id.* § 2.01(1) (requiring voluntary act); *id.* § 2.02(1) (requiring culpability).

Finally, the MPC attempted to solve the subjective/objective problem by defining purpose, knowledge, and recklessness as subjective inquiries, but defining negligence as an objective one.[34]

Each specific crime definition contains a required mental state. Thus, for example, murder is defined as a purposeful or knowing killing,[35] and manslaughter as a reckless killing.[36] The gist, and genius, of the MPC solution to the culpability discordance was to divide wrongful behaviors based on two dimensions: desires and risk taking. The purposeful act is purely desire-based. An actor acts purposefully if he desires the very result caused by his wrong. Knowing is the conscious willingness to take a "practically certain" risk of harm to accomplish some other desire. Reckless is the conscious willingness to take a somewhat lower risk of harm ("substantial and unjustifiable risk") to accomplish some other desire. And negligence is taking but being unaware of a substantial risk of harm.

Although there were many detractors,[37] the MPC formulation of culpability was hailed by most commentators as a reasonable attempt

---

[34] *See* John L. Diamond, *The Myth of Morality and Fault in Criminal Law Doctrine*, 34 AM. CRIM. L. REV. 111, 123 n.73 (1996) ("[The MPC] defines negligence in objective terms, as contrasted with recklessness where subjective awareness is required."). The subjective/objective controversy has nevertheless remained heated in the general context of justification versus excuse. *See, e.g.*, Kent Greenawalt, *The Perplexing Borders of Justification and Excuse*, 84 COLUM. L. REV. 1897, 1915–18 (1984) (discussing justifications as objective and excuses as subjective). By contrast, the MPC did not attempt to solve the problem of how far into the elemental chain the state of mind requirement runs, only indicating that it runs to each "material element" of the crime. MODEL PENAL CODE § 2.02(1); *see also* Robinson & Grall, *supra* note 1, at 691–99 (1983) (discussing MPC method of defining culpability terms "in relation to each objective element of an offense").

[35] MODEL PENAL CODE § 210.2(1)(a). Unlike the common law, the MPC did not distinguish first degree murder (requiring a purposeful killing *after deliberation*) from second degree murder (requiring only a purposeful or knowing killing). That is, the MPC followed the English model in this regard. *See supra* note 21 (discussing English law of homicide).

[36] MODEL PENAL CODE § 210.3(1)(a). Manslaughter is alternatively defined as a purposeful or knowing killing if committed in the heat of passion. *Id.* § 210.3(1)(b).

[37] With regard to the MPC's responsibility conditions, most critics fell into what we will call the "over-determined" school, arguing that one or more of the formulations conceptually and/or practically bled into neighboring ones, at least in some kinds of cases. *See, e.g.*, Kathleen F. Brickey, *The Rhetoric of Environmental Crime: Culpability, Discretion, and Structural Reform*, 84 IOWA L. REV. 115, 122 (1998) (purposeful = knowing); Michael T. Cahill, *Attempt, Reckless Homicide, and the Design of Criminal Law*, 78 U. COLO. L. REV. 879, 902 (2007) (knowing = reckless); *cf.* MODEL PENAL CODE § 2.08(2) (negligent = reckless in cases of self-induced intoxication). Other critics question whether the MPC missed the boat entirely by talking about a criminal's mental state as though such a mental state were a real, let alone discoverable, condition. *See, e.g.*, RICHARD A. POSNER, THE PROBLEMS OF JURISPRUDENCE 168 (1990) ("[M]aybe there is nothing to read [in the minds of criminals], or maybe we are not interested in what the murderer was thinking when he pulled the trigger."); Bruce Ledewitz, *Mr. Carroll's Mental State or What Is Meant by Intent*, 38 AM. CRIM. L. REV. 71, 102–07 (2001) (arguing that the fiction of subjective states of mind should be replaced with a robust presumption of intentionality). These debates are

to impose some predictable structure on a notoriously unpredictable and discordant area of the law. State legislatures were even more accepting. By 1983—just 25 years after its promulgation—36 states had largely jettisoned their criminal codes in favor of the MPC.[38] Even in the handful of states that have not adopted it in whole or in part as legislation, the MPC has still managed to find its way into the common law of those states because judges often turn to it for guidance.[39] The MPC is now taught in virtually every law school, with one professor calling it "the principal text in criminal law teaching."[40] Whether in actual legislation, common law, or simply norms accepted by lawyers and judges, the MPC has become "a standard part of the furniture of the criminal law."[41]

What makes this furniture so comfortable, at least as regards culpability, are two central assumptions: (1) These four levels of culpability accurately reflect our moral intuitions about blameworthiness (that is, harm being equal, purposeful behavior is more blameworthy than knowing, knowing more blameworthy than reckless, etc); and (2) jurors, when called upon to do so, will be able to detect the differences between these defined categories. In the experiments we conducted and report on in this Article, we tested both assumptions. Before turning to those experiments, the next Part briefly surveys the few empirical studies to explore these questions.

II
Previous Empirical Studies

Whether jurors are capable of consistently and appropriately distinguishing between the MPC's categories of mental states is an empirical question that legal scholarship has generally ignored.[42]

not so much about the MPC itself as they are a continuation of the debates inside the ALI during its formulation of the MPC, though some scholars have complained that the MPC effectively ended these debates about the nature and deep structures of responsibility. *See, e.g.*, George P. Fletcher, *The Fall and Rise of Criminal Theory*, 1 Buff. Crim. L. Rev. 275, 278 (1998) ("The Model Penal Code ceased being a stimulus to new legislation and became instead a dogmatic resource for teaching criminal law.").

[38] Robinson & Grall, *supra* note 1, at 691–92. The MPC nose counting is complicated by the extent to which some states have adopted it with changes. Depending on the extent of those changes, some states are counted by some commentators as having adopted the MPC in whole, in part, or only being "influenced" by it. *Id.* at 692 n.45.

[39] 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.").

[40] Sanford H. Kadish, *The Model Penal Code's Historical Antecedents*, 19 Rutgers L.J. 521, 521 (1988).

[41] *Id.*

[42] To be sure, some commentators, such as Professor Kevin Jon Heller, have previously asked the question. Heller reflects:

What we currently know about jurors' ability to discern criminal mental states must be almost entirely imported from research on moral reasoning generally. This more general literature is not specifically tailored to the intricacies of the criminal law and thus provides only limited courtroom-relevant insights.

Research by social and moral psychologists, experimental philosophers, and now neuroscientists has provided a great deal of insight about humans' *general* ability to assess the mental states of others.[43] In particular, we have learned much about our ability to distinguish between intentional and non-intentional action—the basic culpability slicing that has been with us for ages.[44] But although we are naturally able to categorize some kinds of mental states,[45] the relevant question for the criminal law is more specific: Are we able, either naturally or

---

> [C]ontemporary criminal law requires jurors to be latter-day Kreskins—to not only reliably distinguish nearly-indistinguishable mental states, but also to accurately determine which of many possible mental states the defendant actually possessed at the time of the crime. Is such mindreading possible? . . . Given the centrality of mens rea to criminal responsibility, we would expect legal scholars to have provided a persuasive answer to this question. Unfortunately, nothing could be further from the truth.

Kevin Jon Heller, *The Cognitive Psychology of Mens Rea*, 99 J. CRIM. L. & CRIMINOLOGY 317, 320–21 (2009); *see also* Justin D. Levinson, *Mentally Misguided: How State of Mind Inquiries Ignore Psychological Reality and Overlook Cultural Differences*, 49 HOW. L.J. 1, 3 (2005) ("Scholars have not yet fully . . . empirically examined the psychological mechanisms involved in understanding others' minds in the legal setting.").

[43] Much of this research falls within a broad "theory of mind" line of research. *See generally* WILLIAM BECHTEL, PHILOSOPHY OF MIND: AN OVERVIEW FOR COGNITIVE SCIENCE (1988) (providing concise overview of philosophy of mind); UNDERSTANDING OTHER MINDS: PERSPECTIVES FROM DEVELOPMENTAL COGNITIVE NEUROSCIENCE (Simon Baron-Cohen, Helen Tager-Flusberg & Donald J. Cohen eds., 2000) (presenting a number of diverse disciplinary viewpoints on the problem of theory of mind).

[44] *See generally* INTENTIONS AND INTENTIONALITY: FOUNDATIONS OF SOCIAL COGNITION (Bertram F. Malle, Louis J. Moses & Dare A. Baldwin eds., 2001) (providing a range of interdisciplinary theoretical and empirical perspectives on how humans understand and explain the actions of others); Bertram F. Malle & Joshua Knobe, *The Folk Concept of Intentionality*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 101 (1997) (presenting empirical evidence of shared folk concept of intentionality). There is evidence that, at least in some contexts, even infants can identify goal-motivated action. *See* John H. Flavell, *Cognitive Development: Children's Knowledge About the Mind*, 50 ANN. REV. PSYCHOL. 21, 35 (1999) (reviewing literature of "evidence that infants come to construe people as agents, that is, as animate beings that, unlike inanimate objects, can move and behave under their own steam"); Amanda L. Woodward, *Infants' Ability To Distinguish Between Purposeful and Nonpurposeful Behaviors*, 22 INFANT BEHAV. & DEV. 145, 157 (1999) (finding that by nine months, infants understand some actions as goal directed).

[45] Whether, and to what extent, we are born "mind readers" is debated in the fields of developmental psychology and neuroscience. *See generally* Simon Baron-Cohen, *How To Build a Baby that Can Read Minds: Cognitive Mechanisms in Mindreading*, *in* THE MALADAPTED MIND: CLASSIC READINGS IN EVOLUTIONARY PSYCHOPATHOLOGY 207, 207 (Simon Baron-Cohen ed., 1997) (discussing evolution of "Mindreading System" that allows individuals to interpret and predict actions of others).

when prompted by instruction, to categorize others' mental states *in the more precise manner required by the criminal law, and by the MPC in particular*?

While experimentalists such as philosopher Joshua Knobe have given us insights about "people's *ordinary* criteria for intentional action,"[46] our goal in this Article is to assess individuals' behavior when they are in the not-so-ordinary position of being called upon as jurors to assess the culpability of a defendant using the precise set of MPC definitions. On this count, while empirical findings in psychology and philosophy can be translated into legally relevant categories, the translation is difficult.[47] For instance, the psychology literature does not typically use the MPC categories of "knowingly," "recklessly," and "negligently." It instead examines similar, but not wholly analogous, concepts such as "belief," "desire," "awareness," and "foreseeability."[48]

Still, there has been a small body of legally relevant research on the assessment of culpability. Nearly twenty years ago, attorney Laurence Severance teamed up with psychologists Jane Goodman and Elizabeth Loftus to conduct a study with forty-six undergraduates at the University of Washington.[49] The researchers wanted to see how the students would interpret and apply the legal definition of four culpable mental states: intent, knowledge, recklessness, and negligence.

---

[46] Julia Kobick & Joshua Knobe, *Interpreting Intent: How Research on Folk Judgments of Intentionality Can Inform Statutory Analysis*, 75 BROOK. L. REV. 409, 420 (2009) (emphasis added); *see also* Knobe, *supra* note 15, at 193 (finding that when evaluating vignettes, subjects "seem considerably more willing to say that a side-effect was brought about intentionally when they regard that side-effect as bad than when they regard it as good").

[47] *See generally* Bertram F. Malle & Sarah E. Nelson, *Judging* Mens Rea*: The Tension Between Folk Concepts and Legal Concepts of Intentionality*, 21 BEHAV. SCI. & L. 563, 578 (2003) ("[J]urors' folk concepts [of intentionality] clash with the legal concepts that they are expected to apply."); Daniel McGillis, *Attribution and the Law: Convergences Between Legal and Psychological Concepts*, 2 LAW & HUM. BEHAV. 289, 297 (1978) ("Many of the legal discussions of attributional issues include highly subtle distinctions among concepts that have not yet appeared in social science theorizing on attributional judgment.").

[48] *See, e.g.*, Malle & Nelson, *supra* note 47, at 567 (using these terms to describe the "folk concept of intentionality"). Because we draw on multiple disciplines in this Article, our literature review spanned beyond the traditional legal sources indexed in Westlaw's Journal and Law Reviews (JLR) database. We also looked for relevant work in PsychInfo, ISI Web of Knowledge, pubMed databases, as well as in specific journals (such as *Law and Human Behavior*, *Behavioral Sciences and the Law*, *Journal of Social Psychology*, and *Journal of Experimental Social Psychology*). Having performed this extensive search, we believe that we have identified all studies directly on point.

[49] Laurence J. Severance, Jane Goodman & Elizabeth F. Loftus, *Inferring the Criminal Mind: Toward a Bridge Between Legal Doctrine and Psychological Understanding*, 20 J. CRIM. JUST. 107, 109 (1992). Surprisingly, this study has to date been cited only once within the Westlaw JLR database.

They hypothesized, similar to our expectations, that "to the extent that jurors' assumptions or predispositions do not match the distinctions made by law, jurors will experience difficulty in applying legal concepts and may not apply the legal concepts in ways that have been assumed."[50]

In order to test this hypothesis, the researchers randomly assigned students into one of three experimental groups, each of which received a different version of a booklet containing a number of tasks to complete. Subjects in group one (the "Own Definition" group) were asked to define, in their own words, the following terms, appearing in random order: "criminal intent," "criminal knowledge," "criminal recklessness," and "criminal negligence."[51] Subjects in the second group (the "Legal Definition" group) were given full definitions of the four mental states as delineated in the Washington Pattern Jury Instructions for Criminal Cases, which are modeled on the MPC. Subjects in the third group (the "Baseline" group) were not asked to provide their own definitions of any terms, nor were they given any information about the mental states. To measure subjects' ability to apply the legal definitions of mens rea in specific factual contexts, the experimenters presented subjects in all groups with three scenarios. Each scenario consisted of a core "stem" describing a situation in which one person caused harm to another.[52] Each scenario stem was followed by four alternative descriptions of the manner in which the incident occurred, corresponding to the four distinct legal categories of mens rea. Subjects were asked to rank the four explanations by indicating how much punishment they would assign to each on a scale from one to four, with one indicating the most punishment and four indicating the least punishment.

This experimental design allowed the researchers to compare the effects of (1) a baseline condition (no instructions and no prompt) vs. (2) providing jury instructions and (3) prompting thought about a subject's own notions of mens rea. The researchers hypothesized that subjects would not be capable of making refined mens rea distinctions, and thus would not rank order the four variations in the same order as the criminal code. They also hypothesized that exposure to the

---

[50] *Id.* at 108.

[51] *Id.* at 109.

[52] An example of one of the three scenario stems used is:

> A group of high school students is leaving a football game very agitated by their team's loss to their cross-town rival. When they see a group of students from the other school, one person tosses a bottle into the air. The bottle strikes the ground and flying glass cuts several people.

Id. at 110.

Washington Pattern Jury Instructions would improve a subject's ability to rank order the mental states according to the degree of culpability involved.

Severance et al. found that while their intuitions about juror confusion were accurate, their hypothesis about the value of instructions was not supported.[53] Of the four mental states examined—intent (I), knowledge (K), recklessness (R), and negligence (N)—the only distinctions that subjects could make were between the extremes of I and N. In the middle—I vs. K; I vs. R; K vs. R; K vs. N; R vs. N—subjects were not able to reliably make distinctions.[54] Jury instructions made no difference in subjects' ability to make these distinctions. When assigning punishment levels, subjects were similarly able to differentiate only between the extremes of intentional and negligent acts, and not between any of the more fine-grained distinctions. Contrary to expectations, this was true both for those subjects who did not have the legal definitions provided and for those who did.[55]

Whereas Severance et al. conducted just one study, in the early 1990s legal scholar Paul Robinson and psychologist John Darley ran a series of experiments to determine the amount of liability and punishment individuals assign when evaluating different levels of culpability for various selected offenses.[56] In several of their studies, the experimenters sought to compare the MPC's treatment of different culpability levels with the natural intuitions of the community. Subjects were presented with six scenarios containing instances of nonconsensual sexual intercourse, statutory rape, and property damage offenses involving damage to either a house or to unimproved property. Each scenario had four variations, allowing the scenario actor's level of culpability to vary among knowledge, recklessness, negligence, and faultlessness. The researchers designed scenarios with the same basic fact patterns, allowing the four variations in mental states to be clearly

---

[53] *Id.* at 115.

[54] The researchers found that when rank ordering mental states, "legally naive subjects could not, on their own, reliably agree on differentiation between 'criminal knowledge' and 'criminal recklessness' nor reliably distinguish these from other legally relevant mental states." *Id.*

[55] In addition, Severance et al. carried out a content analysis of subject-generated mental state definitions. They sought to determine, qualitatively, the extent to which subjects' definitions of the mens rea terms varied from the legal definitions. The researchers found that subjects often had their own set of preconceptions that deviated from the legal concepts of mens rea. *Id.* at 114.

[56] PAUL H. ROBINSON & JOHN M. DARLEY, JUSTICE, LIABILITY, AND BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW (1995). The book reports eighteen studies, which were designed and executed in seminars at Rutgers University School of Law in Fall 1990 and Spring 1991. *Id.* at xv.

communicated.[57] The four variations of each scenario were given together, in reverse order of culpability level (faultlessness, negligence, recklessness, and knowledge), and the subjects then assigned liability.[58]

In contrast to the Severance et al. study, which found that individuals did *not* categorize mental states in the way the law presumed they did, the results from Robinson and Darley's experiments suggest that subjects' assignment of liability and punishment *are* generally in accord with the MPC. Within each of the six scenarios, the level of liability and punishment assigned increased as the manipulated level of culpability increased.[59] Robinson and Darley's experimental results give us reason to think that, in some circumstances, individuals' mental state evaluations are aligned with the MPC mental state hierarchy. Indeed, Robinson and Darley conclude that the Knowing/Reckless boundary is one at which subjects would likely make liability distinctions.[60]

A decade after the Robinson and Darley study, law professor Justin Levinson conducted an experiment that explored the mediating role of culture in the assessment of defendants' mental states.[61] Levinson compared the responses of undergraduates at Beijing University in China with those of undergraduates at Harvard and the University of California, Berkeley. Subjects read one of four vignettes describing a criminal act committed by an actor whose state of mind was ambiguous. Subjects were then asked to identify the defendant's mental state on a seven-point scale of increasing culpability. Levinson found, for three of the four fact patterns utilized, that the responses of both the American and Chinese undergraduates did not match those predicted by the MPC.[62] Levinson also found differences between the

[57] For example, in the case of property damage, subjects were told, in the faultless condition, that the actor had been informed by his lawyer that the property was his. In the corresponding negligence condition, subjects were told that the actor had not realized that the title of the property had not yet transferred to him, but a reasonable person would have realized this. *Id.* at 86, 221.

[58] Subjects were asked to assign liability on a scale from 0–11, with 0 corresponding to liability but no punishment, 11 corresponding to death, and gradations in between. The liability-punishment scale also included the option of N, which corresponded to no criminal liability. *Id.* at 223.

[59] *Id.* at 87–90.

[60] *Id.* at 87 ("The responses of our subjects, if modeled in the code, would assign a higher degree of liability to the knowing versus the reckless commission of all offenses.").

[61] Levinson, *supra* note 42.

[62] Only when averaging over all four fact patterns does Levinson find some evidence that "participants maintained a folk mental state hierarchy," placing "purpose above knowledge above recklessness" in their punishment ratings. *Id.* at 20. But these results were not robust, as they did not hold in each fact pattern when analyzed individually. *Id.* at 21.

American and Chinese responses, with the Chinese students choosing more culpable states of minds, on average, than the Americans.[63] Chinese students were also more likely to convict for attempted murder, and for assault and battery.[64] These findings remind us of the importance of cultural variation in mens rea evaluations.

Taken together, the studies by Severance et al., Robinson and Darley, and Levinson paint an incomplete, and at times contradictory, picture of the ability of jurors to evaluate criminal defendants' mental states. This may be due in part to two methodological shortcomings: (1) experimental subjects were unrealistically exposed to repeated variations of the same fact patterns; and (2) some of the subject pools in the experiments relied heavily on college and graduate students, and were thus not representative of the general population.

In an actual criminal trial, jurors are exposed to one fact pattern, albeit presented and interpreted differently by prosecution and defense.[65] In both the Severance et al. and Robinson and Darley studies, however, subjects saw all mental state variations of the same underlying fact pattern. That is, subjects had an opportunity to read about John acting purposefully, then about John acting knowingly, then about John acting recklessly, and then about John acting negligently.[66] This is described as a "within-subjects" design in psychology research because the variable of interest—John's mental state—is being varied *within* a given subject's treatment. A concern with such designs, readily acknowledged and discussed by Robinson and Darley, is that "[b]ecause subjects see all of the test scenarios [that is, John in all of his mental state variations] and because the scenarios differ only in relation to [the mental state], the differences that are being tested inevitably are obvious to the subjects."[67] In other words, subjects in these experiments may be paying more attention to differences between mental states than they ordinarily would in a real-world legal context.

---

[63] *Id.* at 22.

[64] *Id.*

[65] Levinson recognized this problem in his study. *Id.* at 27. Of course, after the prosecution presents the jury with one set of facts, the defense might present it with evidence of alternative facts suggesting a less culpable mental state. This sometimes happens, but more often than not it is counsel's argument that suggests differing mental states, rather than conflicting evidence about the act itself or the circumstances of its commission.

[66] Robinson and Darley concede that it would have been better methodologically to randomize the order of the variations. ROBINSON & DARLEY, *supra* note 56, at 288. However, even this modification may not eliminate the bias. A better approach is to have a sufficient number of scenarios such that subjects see a certain scenario only once, and make one rather than four mens rea judgments for each scenario.

[67] *Id.* at 222.

If, for example, subjects had not read about John acting purposefully, knowingly, and recklessly, would they have made the same judgments about him acting negligently? Absent the exposure to the other mental state scenarios, would they have even recognized that this was a negligent mental state? Research designs that expose subjects to the same fact pattern multiple times cannot answer these questions. Our experiments improve upon these earlier research designs by exposing subjects to only one scenario from each fact pattern.

In addition to the multiple exposure problem, the generalizability of previous findings is limited by the nature of their subject samples. All three of the previous studies relied on convenience samples for their subject pool. As the name implies, convenience samples are samples of subjects identified out of convenience, such as students on the campus of the university where the researcher works. In the Severance et al. and Levinson studies, college students comprised the sample. The Robinson and Darley sample was somewhat more diverse, but still generated by convenience.[68] For practical reasons such convenience samples are the norm in psychology research, but the findings may not generalize to a jury pool that is significantly more diverse in age, education, and geography.[69]

To be sure, American juror pools include college students. But they are comprised primarily of individuals older than twenty-two, and also include the large percentage of Americans who do not hold a

---

[68] The authors are upfront about this issue when they write that a "difficulty with our studies" is "the particular procedures we used for selecting our respondents. Putting it inelegantly, we grabbed whomever we could get our hands on. Typically, the subjects were neighbors, family, or friends of the students." *Id.* at 222.

[69] The effect on the validity of experimental findings of heavy reliance on undergraduate subjects has been much discussed. For one critique, see Steven D. Levitt & John A. List, *What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?*, 21 J. ECON. PERSP. 153, 154 (2007) ("[G]reat caution is required when attempting to generalize lab results out of sample . . . ."). There is a great deal of literature on the subject. *See generally* Marc Hooghe et al., *Why Can't a Student Be More Like an Average Person?: Sampling and Attrition Effects in Social Science Field and Laboratory Experiments*, 628 ANNALS AM. ACAD. POL. & SOC. SCI. 85 (2010) (arguing for inclusion of non-student samples in field and laboratory experiments); Robert A. Peterson, *On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-Analysis*, 28 J. CONSUMER RES. 450 (2001) (documenting differences and similarities between college student subjects through second-order meta-analysis). *But see generally* Jerald Greenberg, *The College Sophomore as Guinea Pig: Setting the Record Straight*, 12 ACAD. MGMT. REV. 157 (1987) (attempting to dissuade organizational researchers from prematurely dismissing findings of studies that used student samples). This discussion stretches back over half a century. *See, e.g.*, Maurice L. Farber, *The College Student as Laboratory Animal*, 7 AM. PSYCHOLOGIST 102, 102 (1952) (arguing that college students are desirable laboratory subjects because they are human, easily available, and comparatively alert, responsive, and articulate).

college degree.[70] In order to produce more generalizable findings, a more representative sample is required. Our study addresses this challenge.

III
Our Experiments

A. General Methodological Background

The experimental design for each of our studies required individuals to read short scenarios and to answer a single question about the scenario's protagonist. The first step in experimental design was to develop scenarios that were readily accessible to the subject (that is, straightforward and reasonably believable on their face), clearly communicative of a distinct MPC mental state, and short enough that subjects could read multiple scenarios within a reasonable amount of time.[71] Moreover, because previous research has pointed to the interaction of harm level with mental state determinations, we also aimed to vary the harm level across our scenarios.[72]

---

[70] Based on census data from 2000, the U.S. Census Bureau reports that twenty-four percent of Americans age twenty-five and older had completed a college degree. Kurt J. Bauman & Nikki L. Graf, U.S. Census Bureau, Educational Attainment: 2000, at 1 (2003). Jury composition of course varies, by design, across communities. In the abstract, it is therefore impossible to specify what a "typical" jury looks like; nevertheless, a 2008 study from the State of Washington, using surveys to evaluate the demographics of jurors in three Washington state counties, provides useful data. Wash. State Ctr. for Court Research, Juror Research Project: Report to the Washington State Legislature 4 (2008). The study compared juror demographics with county census data. The study found that jurors were on average older than the general population in the county. *See id.* at app. D (presenting demographic characteristics of jurors for each study site). Only one percent of the jurors reported being students. *Id.*

[71] These constraints raised a number of questions about how to effectively and efficiently communicate the protagonist's motivation and intent. John's action in each of our scenarios was explained to subjects with a simple, and typically neutral, motivation. For instance, in one scenario subjects read that John acted because he was angry after an argument with a player on an opposing softball team. Scenario construction was also sensitive to the fact that moral judgments about the actor involved may influence mental state assessment. *See* Thomas Nadelhoffer, *Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality*, 9 Phil. Explorations 203, 208 (2006) (arguing that, to the extent that moral considerations affect folk ascriptions of intentional action, the ability of a defendant who is being prosecuted for a serious crime to receive a fair and unbiased assessment by jurors is undermined).

[72] *Compare* Joshua Knobe, *The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology*, 130 Phil. Stud. 203, 214 (2006) (arguing that the moral status of an agent's behavior affects the subject observer's judgment as to whether or not an action was performed intentionally), *with* Edouard Machery, *The Folk Concept of Intentional Action: Philosophical and Experimental Issues*, 23 Mind & Language 165, 172–73 (2008) (arguing that Knobe draws no principled distinction between what constitutes subject competence with a given concept and what results merely from factors that affect one's use of the concept on any given occasion).

Applying these principles, we drafted scenarios featuring a protagonist (always named John) whose actions cause differing levels of harm. We organized the specific scenarios within "themes." We use the term "theme," which is akin to previous researchers' term "stem," to refer to the general fact pattern (for example, "John drops wood planks onto a bike trail, and two bikers crash as a result."). We use the term "scenario" to refer to a fact pattern with a specific mental state (for example, "While carrying wood planks, John drops some onto the trail and doesn't pick them up because he wants to start the carpentry, even though he is aware that there is a substantial risk that bikers will hit the planks and be injured.").[73] Thus, within each theme there are five scenarios: one each for purposeful, knowing, reckless, negligent, and blameless.

We created thirty themes, ten in each of three harm-level categories: high harm (causing death or serious injury), medium harm (causing minor injury or great property damage), and low harm (causing no injury or minor property damage). Within each of these thirty themes, we constructed one scenario for each of the four MPC mental states plus one non-culpable mental state: (1) purposeful, (2) knowing, (3) reckless, (4) negligent, and (5) blameless. Thus, we wrote a total of 150 unique scenarios.[74]

Each scenario was comprised of exactly three sentences. Within a given theme (that is, a general fact pattern), the first and third sentences were identical.[75] Holding the first and third sentences constant allowed us to attribute resulting behavioral differences in punishment ratings between same-themed scenarios to changes in the second sentence, which in turn enabled inferences about John's mental state. Mental state signals were rotated systematically across themes so that each signal was used exactly six times. Mental state signals were also counterbalanced evenly within and across all of the three harm levels. Table 1 illustrates, for one of these thirty themes, how each scenario was constructed.

---

[73] All scenarios were constructed so that they would have roughly the same total number of words. Scenario length was 73 words, +/- 2 words.

[74] For the full set of scenarios, see Appendix B.

[75] There were just a few exceptions where we needed to change the word "the" to the word "a."

TABLE 1: ILLUSTRATION – VARYING MENTAL STATE WITHIN A SINGLE THEME

Sentence 1 of 3	Mental State in Scenario	Sentence 2 of 3	Sentence 3 of 3
John is doing carpentry work on his house, which abuts a public mountain bike trail.	Purposeful	Angry at the mountain bikers for making too much noise when biking past his house, one day while carrying a large armload of planks, John desires to injure some bikers and drops some of the planks on to the bike trail.	Two bikers passing by at that moment hit the planks, crash as a result, and are seriously injured.
	Knowing	While carrying wood planks, John drops some onto the trail and doesn't pick them up because he wants to start the carpentry work, even though he is practically certain that in doing so bikers will hit the planks and be injured.	
	Reckless	While carrying wood planks, John drops some onto the trail and doesn't pick them up because he wants to start the carpentry, even though he is aware that there is a substantial risk that bikers will hit the planks and be injured.	
	Negligent	One day while John is carrying wood planks from his shed to his workshop in order to begin building a new set of steps for his house, he drops some of the wood planks onto the bike trail without noticing.	
	Blameless	One day while John is carefully carrying wood planks from his shed to the backyard where he is building a wood porch, a sudden strong gust of wind causes John to inadvertently drop several planks, despite his best efforts not to.	

Signaling mental states is, of course, a function of not just word choice but also of scenario context. A particularly vexing challenge for scenario construction was how to structure the scenarios such that we could clearly communicate John's mental state with regard to the harm being caused. We addressed this challenge in the following way.

If we consider *y* as the harm variable, and consider *x* as the variable for John's action in the scenario, then within each theme *x* varies, *y* remains constant, and the general relationship between *x* and *y* is:

Purposeful: John decides to cause (or bring about) *y* via *x*

Knowing: John does *x*, practically certain that it will result in *y*

Reckless: John does *x*, aware there is a substantial risk that *y* will occur

Negligent: John does *x* carelessly, thus causing *y*

Blameless: John does *x*, and despite being as careful as he could be, he accidentally causes *y*

But using only these MPC signaling terms creates a problem with habituation. An experiment that exposes subjects to identical signaling language for each mental state risks becoming just a reading test. So we devised five alternative phrases to signal each mental state. The language we developed is reported in full in Table 2.[76]

TABLE 2: LANGUAGE USED TO SIGNAL JOHN'S MENTAL STATE IN SCENARIOS

Note: For each of the five mental state categories, we systematically rotated between five different signaling phrases, in order to prevent subjects from identifying a mental state purely on the phrase employed.

1) Purposefully (consciously intends the specific harm)
   a. Decides to
   b. Intends that (or with the intention of)
   c. Desires that
   d. Wants to
   e. Chooses to

[76] The Model Penal Code uses the phrase "acts purposely" and not the word purposeful. In courts, different constructions of the phrase, including "purposefully," are utilized. *See, e.g.*, Maxwell v. State, 41 S.W.3d 402, 408 (2001) ("A person acts purposefully with respect to her conduct when it is her conscious object to engage in conduct of that nature or to cause such a result."). Because the precise phrasing used to describe the "purposely" mental state varies, some states explicitly note that equivalent terms have the same meaning. *See, e.g.*, ARK. CODE ANN. § 5-1-102(17) (2009); N.J. STAT. ANN. § 2C:2-2(b)(1) (West 2005). In this Article, we use the terms "purposely" and "purposefully" interchangeably.

TABLE 2 (CONT'D)

2) Knowingly (similar language as Purposefully, but with contextual clarification that John doesn't separately *intend* the harm that occurs; he is instead aware that acting to fulfill his separate intention *will certainly cause* (100% certain) the harm that does happen)
   a. Practically certain that [the harm will occur]
   b. Aware that [the harm] will almost certainly occur
   c. Almost positive that [the harm will occur]
   d. Virtually certain that [the harm will occur]
   e. Understands that [the harm] is almost guaranteed to occur

3) Recklessly (very heavily discounts or disregards the risk)
   a. Aware there is a substantial risk [the harm might occur], but chooses to ignore it.
   b. Realizes it is very likely [the harm might occur], but decides to act anyway
   c. Conscious of the likelihood [of the harm], but simply doesn't care
   d. Understands [that the harm could easily happen], but decides to risk it.
   e. Knows there is a good chance that [the harm will occur], but chooses to act anyway.

4) Negligently (objective risk flagged in scenario; emphasis on subjective ignorance of risk)
   a. Carelessly
   b. Wasn't paying attention
   c. Hurriedly (made clear through context)
   d. Without even noticing
   e. Overlooks

5) Blamelessly
   a. Despite being as careful as he could, accidentally [causes harm]
   b. [Act is involuntary]
   c. Unavoidably [causes harm]
   d. Through an honest mistake [causes harm]
   e. Inadvertently [causes harm] despite his best efforts.

Recognizing the importance of pre-testing the utility of the signaling language before using it in our experiments, we turned to nine criminal law professors for external validation that our scenarios, and in particular our alternative sets of mental state signaling language, were in fact communicating the mental state we posited they did. Each professor read all 150 scenarios, presented in a random order, and assigned a mental state. The law professors were able to sort these with an 84% accuracy rate (rising above 90% if one outlier is

removed). This bolsters our confidence that we were indeed signaling the mens rea categories as defined by the MPC.

We also ran a preliminary study to validate that our assignment of low, medium, and high harm levels corresponded to subjects' perception of the harm level.[77] The results of that preliminary study confirmed our assumptions about the level of harm in each scenario, and our groupings of those levels into the three categories of low, medium, and high. We used standardized harm ratings from this initial study in our subsequent models to control for the potential confounding effect of theme harm levels, and to investigate whether sorting ability varies by harm level (that is, do subjects more accurately sort mental states when they evaluate scenarios in which John commits more harm?).[78]

B. Experimental Design

In order to know what value, if any, might be added by including the jury instructions providing the MPC definitions, we had to establish a baseline model in which participants were given no MPC definitions to guide their rating (Experiment 1: "How Do Subjects Punish with No MPC Instructions?"). Participants in Experiment 1 were simply presented with a fact pattern and asked to make a punishment rating.

We then turned to more realistic, courtroom-like situations of asking subjects to make punishment judgments after they read the MPC definitions once (Experiment 2: "How Do Subjects Punish After Reading the MPC Definitions Once?"); and asking subjects to make punishment judgments while having continuous access to the definitions (Experiment 3: "How Do Subjects Punish When They Have Continuous Access to the MPC Definitions?"). We used the Colorado

---

[77] In order to validate our harm level groupings of low harm (themes 1–10), medium harm (themes 11–20), and high harm (themes 21–30), an independent sample of fifty subjects was randomly presented with each of the thirty blameless scenarios. After reading each harm description, subjects were asked: "On a scale from 0–9, with 0 being no harm and 9 being maximum harm, how harmful is <theme-specific description of harm>? (For example, "How harmful is having coffee spilled on completely worthless junk mail?"). On the 0–9 harm rating scale, the average rating for low harm scenarios was 1.6, the average for medium harm scenarios was 4.5, and the average for high harm scenarios was 7.1. Post-estimation chi-squared tests confirmed that the high harm themes were significantly rated as more harmful than medium themes, $F(1, 27) = 40.11$, $p < .001$, and that the medium were rated more harmful than low, $F(1, 27) = 54.81$, $p < .001$.

[78] Subject harm ratings were standardized within subject, to account for variance due to subject-specific factors. The standardized harm rating scores discussed *infra* Section III.C.1 can be understood as the subject's harm rating as measured by the number of standard deviations above/below the subject's mean score for all thirty scenarios.

Pattern Jury Instructions for our culpability definitions.[79] These instructions are reproduced in Table 3.

TABLE 3: MENTAL STATE DEFINITIONS USED IN EXPERIMENTS 2–5

A crime is committed when the defendant has committed a voluntary act prohibited by law accompanied by a culpable mental state. Voluntary act means an act performed consciously as a result of effort or determination. Culpable mental state means either purposefully, knowingly, recklessly or negligently, as explained in this instruction. Proof of the commission of the act alone is not sufficient to prove that the defendant had the required culpable mental state. The culpable mental state is as much an element of the crime as the act itself.

1. Purposefully: A person acts "purposefully" when his conscious objective is to cause the specific result.

2. Knowingly: A person acts "knowingly" when he is aware that his conduct is practically certain to cause the result.

3. Recklessly: A person acts "recklessly" when he consciously disregards a substantial and unjustified risk that a result will occur or that a circumstance exists.

4. Negligently: A person acts "negligently" when, through a gross deviation from the standard of care that a reasonable person would exercise, he fails to perceive a substantial and unjustified risk that a result will occur or that a circumstance exists.

5. Blamelessly: A person is "blameless" even though he may have caused harm, if he lacked any of the culpable mental states defined above.

In Experiment 2, we provided the MPC definitions just once, at the start of the experiment, and told subjects: "We encourage you to keep these five mental states in mind and to use the full range of the rating scale (ranging from 0 to 9, with 0 being no punishment and 9 being extreme punishment) for both the practice and experimental scenarios." In Experiment 3, instead of offering the MPC definitions just once, we made the definitions available on the bottom of the computer screen throughout the experiment. Manipulating access to the MPC definitions is consistent with variations across jurisdictions in the access jurors have to mental state definitions.[80] This combination of

[79] We chose the Colorado Pattern Jury Instructions because the judge co-author of this paper, who presides in the Second Judicial District (Denver) in the State of Colorado, is very familiar with both the language of the instructions and how they are used in practice.

[80] In criminal trials, the mental state definitions, along with all the other elements of the charged crime, are presented to the jury by the judge as part of the written jury instructions. In the past, judges read the jury instructions to the jury (a common law remnant of the days when jurors were illiterate), but by tradition most federal judges did not provide

experiments allowed us to see if varying the delivery of instructions produces different patterns of punishment.

The first three experiments told us much about how individuals punish, but left open the question of mechanisms—how had subjects arrived at a given punishment rating? In particular, we wanted to tease out the distinction between (a) subjects who saw differences in mental states, but punished the same for both; and (b) subjects who simply saw no difference in mental states to begin with. To distinguish between the two, we developed a fourth experiment to determine how well subjects were able to correctly identify each MPC category (Experiment 4: "Can Subjects Distinguish Between Mental States?"). Finally, in order to see if practice with the instructions makes for (more) perfect punishment alignment, we designed an experiment in which subjects completed a mental states sorting task, and then made their punishment ratings (Experiment 5: "How Do Subjects Punish After They Have Practiced Sorting Mental States?").

As noted in Part II, one problem with previous studies is that they exposed subjects to multiple mental states within the same fact pattern.[81] To avoid this problem, in each of our five experiments our subjects read only thirty of the 150 short scenarios, six from each of the five mental states (purposeful, knowing, reckless, negligent, and blameless). Subjects were randomly assigned one scenario from each of the thirty themes.[82] In the rating experiments, after reading each scenario subjects were asked: "On a scale from 0–9, with 0 being no punishment and 9 being extreme punishment, how much should John be punished for his behavior?"[83] Although we report the raw 0–9

---

the jury with a written copy. Spurred in part by various jury reform efforts, the current trend, even in federal courts, is to provide each individual juror with a copy of the instructions.

[81] *See supra* note 66 and accompanying text (describing this multiple exposure problem).

[82] Subjects also were given five practice scenarios, one from each mental state, before the actual experiment, in order to familiarize them with the interface and the experimental task. These practice themes were developed in addition to the thirty themes used in the actual experiment.

[83] Research in moral psychology has found that individuals may assign blame differently than they assign punishment. *See, e.g.*, Jennifer K. Robbennolt, *Outcome Severity and Judgments of "Responsibility": A Meta-Analytic Review*, 30 J. APPLIED SOC. PSYCHOL. 2575, 2580 (2000) (discussing a variety of outcome variables that researchers have used to measure responsibility judgments). To account for this possibility, we ran a set of *blame* rating experiments, identical to the punishment rating experiments, except for a change in the rating question asked. Thus, we re-ran experiments 1, 2, 3, and 5 with a focus on blame rather than punishment. In these additional experiments, subjects were asked, after reading each scenario: "On a scale from 0–9, with 0 being not at all blameworthy and 9 being extremely blameworthy, how blameworthy is John for his behavior?" As reported in more detail in Appendix A, the results from the blame rating experiments followed the

punishment scores throughout this Article, we also conducted additional analysis to alleviate concerns about inter-subject subjectivity.[84]

We conducted the experiments in March and May 2010.[85] We used a web-based experimental platform, which allowed us to recruit a large and diverse sample of subjects.[86] We used the web survey firm Qualtrics to recruit subjects from across the country.[87] Research using Qualtrics-based experiments has been published and presented in a number of academic fields, which suggests that it meets scholarly expectations for quality online web-based experiments.[88]

All subjects recruited by Qualtrics were United States citizens, ages eighteen to sixty-five. In order to better approximate a jury-eligible subject pool, we filtered out, via an initial screening question, subjects who indicated that they had been convicted of a felony. Qualtrics recruited subjects through opt-in survey panels drawn from the general population. The number of subjects for each experiment, reported in Table 4, allowed enough statistical power to robustly test our hypotheses. For the baseline Experiment 1 ("How Do Subjects

---

same pattern as the punishment rating experiments we discuss in the text. Thus, we are confident that our results are not an artifact of asking about punishment instead of blame.

[84] We employed "standardized" punishment ratings to account for the fact that individuals may interpret the punishment rating scale differently. The process of standardization was used to alleviate concerns about inter-subject subjectivity in interpreting the punishment scale. For instance, it corrects for the situation where subject A believes a rating of 9 represents the death penalty, but subject B believes that 9 represents only twenty years in prison. Standardization transforms each raw punishment score (the 0–9 rating) into a "standardized" rating (also known as a "z-score") which can be understood as: "For scenario X, how many standard deviations above/below the subject's mean rating (for all thirty scenarios rated) is his punishment rating?" As a practical matter, although there are a few differences in the results using the different measures, the substantive conclusions of our analysis remain the same whether we use the actual punishment scores or the standardized measures. For ease of interpretation, we report only the punishment scores.

[85] All experiments received approval from the University of California, Santa Barbara Institutional Review Board (email granting approval on file with the *New York University Law Review*).

[86] Researchers in psychology have increasingly turned to web-based experiments because they offer a "'large number of participants' and 'high statistical power.'" Ulf-Dietrich Reips, *Standards for Internet-Based Experimenting*, 49 EXPERIMENTAL PSYCHOL. 243, 244 (2002) (quoting Jochen Musch & Ulf-Dietrich Reips, *A Brief History of Web Experimenting*, *in* PSYCHOLOGICAL EXPERIMENTS ON THE INTERNET 70 (M.H. Birnbaum ed., 2000)).

[87] Subject costs were approximately $7 per subject for a nationally representative sample. For more background on the Qualtrics platform, see QUALTRICS: ONLINE SURVEY SOFTWARE, http://www.qualtrics.com (last visited Oct. 6, 2011).

[88] 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*, 73 J. MARKETING 18, 21 (2009); Paul H. Robinson et al., *The Disutility of Injustice*, 85 N.Y.U. L. REV. 1940, 2000 (2010).

Punish with No MPC Instructions?") and Experiment 4 ("Can Subjects Distinguish Between Mental States?") we added additional subjects by running each experiment twice (three months apart, but under identical experimental conditions).[89]

TABLE 4: DISTRIBUTION OF SUBJECTS ACROSS EXPERIMENTS

Experiment	**Number of Subjects**
Primary Experiments: Using Punishment Ratings	
Experiment 1: "How Do Subjects Punish with No MPC Instructions?"	196
Experiment 2: "How Do Subjects Punish After Reading the MPC Definitions Once?"	96
Experiment 3: "How Do Subjects Punish When They Have Continuous Access to the MPC Definitions?"	97
Experiment 4: "Can Subjects Distinguish Between Mental States?"	201
Experiment 5: "How Do Subjects Punish After They Have Practiced Sorting Mental States"	150
Additional Experiments: Using Blame Rating	
Experiment 1b: "How Do Subjects Blame with No MPC Instructions?"	194
Experiment 2b: "How Do Subjects Blame After Reading the MPC Definitions Once?"	96
Experiment 3b: "How Do Subjects Blame When They Have Continuous Access to the MPC Definitions?"	94
Experiment 5b: "How Do Subjects Blame After They Have Practiced Sorting Mental States?"	152
Preliminary Experiment To Assess Harm Levels	50
TOTAL SUBJECTS, ACROSS ALL EXPERIMENTS	*1326*

At the end of the experiment, we collected demographic information from subjects. Table 5 shows these results. While not a truly nationally representative sample, the 1326 subjects who participated in the experiments came from 47 states. Our sample was roughly equal in terms of gender, with 53% of subjects female and 47% male. Our subjects were older, on average, than the comparable U.S.

[89] For both of the experiments, the results from the second round of the experiment were substantively the same as the results from the first round. This reassured us that our results are robust. We report in this Article the results of the analysis of data pooled across both rounds of running the experiment.

population. Our sample was 84% white, higher than the national average. In terms of education, our subjects were slightly skewed toward having more education than the population as a whole. Income distributions of our subjects and the U.S. population as a whole are similar, though not identical. At a minimum, our sample is more reflective of a jury pool, both in its size and demographic makeup, than any of those of the previous studies discussed in Part II.[90]

TABLE 5: DEMOGRAPHICS OF EXPERIMENTAL SUBJECTS (N = 1326)

Education	**Subjects**	**U.S. Census**
Less than High School	1%	18%
High school / GED	21%	30%
Some college	32%	20%
Associate Degree	13%	7%
Bachelor's Degree	22%	17%
Graduate Degree	10%	10%
Income	**Subjects**	**U.S. Census**
< $20,000	19%	$1 to $24,999: 22%
$20,000 - $40,000	31%	$25,000 to $34,999: 19%
$40,000 - $60,000	22%	$35,000 to $49,999: 21%
$60,000 - $80,000	14%	$50,000 to $64,999: 14%
$80,000 - $100,000	7%	$65,000 to $74,999: 6%
> $100,000	7%	$75,000 to $99,999: 8%
Gender	**Subjects**	**U.S. Census**
Male	47%	49%
Female	53%	51%
Age Groups	**Subjects**	**U.S. Census**
18–24	5%	13%
25–34	12%	18%
35–44	19%	19%
45–59	48%	27%
60 +	16%	23%
Race	**Subjects**	**U.S. Census**
White	84%	74%
Non-White	16%	26%
Jury Member (Criminal Case)?	**Subjects**	
Yes	18%	
No	82%	

[90] *See supra* notes 42–70 and accompanying text (discussing sample problems of previous studies).

C. Results

In this Section we summarize the results of our punishment rating experiments. In Appendix A we provide additional discussion of the details of the statistical analysis.

1. *Results from Experiment 1*: *"How Do Subjects Punish with No MPC Instructions?"*

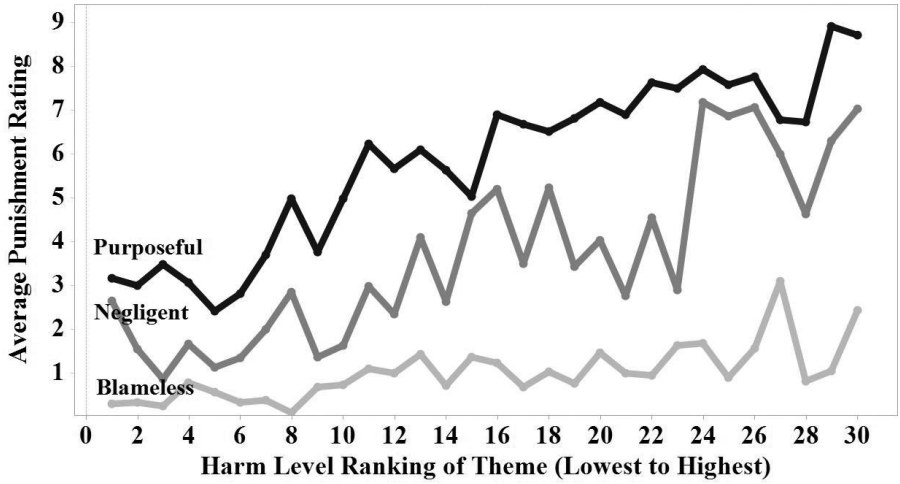
The results from our baseline experiment—assessing subject punishment ratings in the absence of any MPC guidance—suggest that even without exposure to the actual definitions, subjects punish in accordance with MPC guidelines in the purposeful, negligent, and blameless conditions. Figure 1 reports average punishment ratings for purposeful, negligent, and blameless scenarios, plotted by ranking of harm. Punishment ratings were the highest for purposeful action, just as the MPC would predict. At the other end of the spectrum, blameless punishment averages were the lowest, and negligent averages were the second lowest. These results show not only that subjects punished in these categories as the MPC assumes they would, but also that subjects were very good at distinguishing these three categories of mental states from one another (which, of course, is a prerequisite to being able to punish differentially). Note that the marked separation between the lines in Figure 1 persists across the entire range of harm, suggesting that for these three categories of culpability our subjects behaved just as the MPC predicts. That is: (1) they were able to distinguish purposeful (P), negligent (N), and blameless (B) states of mind; (2) where harm is equal, they punished P more than N and N more than B; and (3) as harm increased, the punishment level in both blame categories increased. These differences, confirmed by statistical analysis reported in Appendix A, were statistically significant.

Where the MPC assumptions seem to fail, however, is at the junction between knowing (K) and reckless (R). Figure 2 adds the K and R average punishment rating lines to Figure 1. The intertwining darker (K) and lighter (R) lines in Figure 2 illustrate that K is often punished no differently, or even less harshly, than R, not at all in keeping with the MPC hierarchy assumptions.

To gain a more precise understanding of this K/R boundary line, we re-examined each theme and asked: Is there a statistically significant difference, within this theme, between K and R punishment ratings? A theme-by-theme analysis, presented in detail in Appendix Table A1, reveals that subjects' average punishment ratings for K and R were significantly different for only six of the thirty themes (and in one of these six it was R that was punished significantly more than K).

On one hand, subjects' ability to differentiate punishment in some of the themes serves as a reminder that certain fact patterns may allow for easier distinction between K and R. On the other hand, that we see so much back and forth between the two categories is powerful evidence that our subjects very often did not see the fine-grained distinctions between K and R that the MPC, and many state statutes, presume they do. At least for the scenarios we constructed, it is the exception rather than the rule that subjects can punish K and R as the MPC presumes they will.

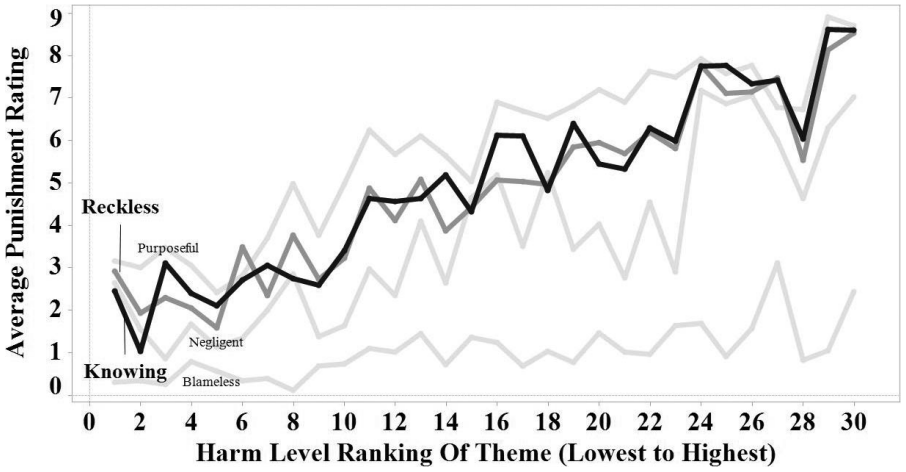
FIGURE 1: AVERAGE PUNISHMENT RATINGS FOR PURPOSEFUL, NEGLIGENT, AND BLAMELESS SCENARIOS FROM EXPERIMENT 1 ("HOW DO SUBJECTS PUNISH WITH NO MPC INSTRUCTIONS?") (PLOTTED BY HARM LEVEL RANKING OF THEME)



What To Notice in Figure 1: The average punishment ratings for purposeful, negligent, and blameless scenarios are completely distinct from one another. That is, they do not cross, and their order is consistent with the assumptions of the Model Penal Code.

Notes: Data for this figure are from Experiment 1 ("How Do Subjects Punish with No MPC Instructions?"). The y-axis plots the average punishment rating for each purposeful, negligent, and blameless scenario in each of thirty themes (averaged across all subjects who rated the particular scenario). Shading indicates the mental state of the scenario.

FIGURE 2: AVERAGE PUNISHMENT RATINGS FOR KNOWING AND RECKLESS SCENARIOS FOR EXPERIMENT 1 ("HOW DO SUBJECTS PUNISH WITH NO MPC INSTRUCTIONS?") (PLOTTED BY HARM LEVEL RANKING OF THEME)



What to Notice in Figure 2: The average punishment ratings for knowing and reckless scenarios cross each other repeatedly, visually presenting what is confirmed by statistical analysis discussed in Appendix A: There is no significant difference between punishment ratings of knowing and reckless scenarios. That is, punishment ratings of K and R actors are *not* consistent with the assumptions of the MPC.

Notes: Data for this figure are from Experiment 1 ("How Do Subjects Punish with No MPC Instructions?"). The y-axis plots the average punishment rating for each knowing and reckless scenario in each of thirty themes (averaged across all subjects who rated the particular scenario). The darker line indicates a knowing mental state, and the lighter line indicates a reckless mental state. The three background lines, which are identical to the lines presented in Figure 1, are (from top to bottom) for purposeful, negligent, and blameless actions.

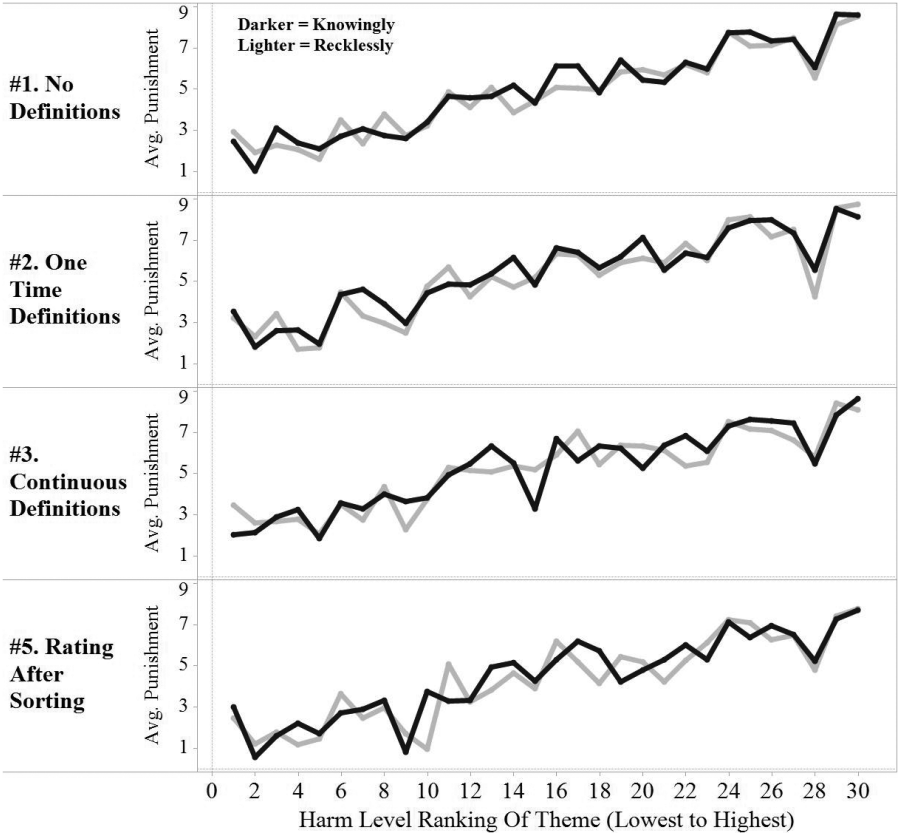
2. *Results from Experiment 2*: *"How Do Subjects Punish After Reading the MPC Definitions Once?" and Experiment 3*: *"How Do Subjects Punish When They Have Continuous Access to the MPC Definitions?"*

The results in Figures 1 and 2 were generated from Experiment 1, in which subjects did not have the benefit of reading the MPC definitions. Theoretically, we might see more differentiated graphs when subjects have the instructions to guide them. This theory, however, finds no support in our results. Even with some nudging—in

Experiment 2 we gave subjects the MPC definitions once at the outset; in Experiment 3 we made the definitions available throughout—the confusion over K and R remains.

Figure 3 presents the average ratings for K and R scenarios, comparing these to the averages from Experiment 1. It is evident in Figure 3 that K and R are frequently muddled together, and that R is often punished (even if not statistically significantly so) at a greater level than K. Statistical analysis, reported in Appendix A, confirms that in both Experiment 2 and Experiment 3, there remains—for the vast majority of the themes—no statistically significant difference between punishment ratings for K and R scenarios. Thus, at least through exposure to the MPC definitions, our subjects cannot be readily trained to sufficiently differentiate their punishment between K and R scenarios.

FIGURE 3: COMPARING AVERAGE PUNISHMENT RATINGS FOR KNOWING AND FOR RECKLESS SCENARIOS, ACROSS FOUR DIFFERENT EXPERIMENTAL DESIGNS



What To Notice in Figure 3: Regardless of the experimental design, the average punishment ratings for knowing and reckless scenarios were very similar, and they frequently reversed. The figure illustrates that regardless of the instructions provided, subjects did not regularly rate knowing behavior as more worthy of punishment than reckless behavior.

Notes: Data for this figure come from Experiments 1, 2, 3, and 5. The y-axis plots the average punishment rating for each knowing and reckless scenario in each of thirty themes, averaged across all subjects who rated the particular scenario. The dark line indicates a knowing mental state, and the light line indicates a reckless mental state.

3. *Results from Experiment 4*: *"Can Subjects Distinguish Between Mental States?"*

There are two distinct, though not mutually exclusive, possibilities for why there is not greater differential in punishment ratings between K and R, even when subjects are given the MPC definitions. First, part of the explanation may be that subjects recognize the correct mental state, but then decide to assign different levels of punishment than the MPC prescribes. For instance, a subject might recognize that an act has been committed recklessly, but see no reason to assign more punishment to the knowing act than he would assign to the same act done recklessly. If this is the case, then the problem is one of rating misalignment between the MPC and subjects' inherent culpability scorecard. A second possible reason for the rating confusion, however, is that the subject cannot, from the start, tell the difference between knowing and reckless acts. In this second possible case, the subject would punish differently *if* he could sort properly. The problem is not one of rating, but of sorting ability. Our final two experiments are designed to see whether either or both of these possible factors contribute to the punishment ratings we see.

Experiment 4 ("Can Subjects Distinguish Between Mental States?") gathered baseline data on subjects' sorting ability. Each subject in Experiment 4 read through 30 scenarios, and in addition to the scenario text, subjects were also provided with the definitions of the mental states. After reading each scenario, subjects were instructed: "Please select from the question options below the definition that best matches John's mental state in this scenario." Experiment 4 allowed us to determine, for each mental state, subjects' ability to correctly decode the mental state signaling language.

The results of Experiment 4, summarized in Table 6, suggest that subjects can identify, with a high degree of accuracy, purposeful and blameless scenarios. Subjects correctly identified purposeful scenarios

78% of the time, and correctly identified blameless scenarios 88% of the time. Subjects were most prone to error in the middle categories of knowing (50% success rate), reckless (40% success rate), and negligent (48% success rate).

Subjects are significantly more likely to correctly identify the purposeful and blameless scenarios. Statistical analysis presented in Appendix A confirms this. But at the same time, both the summary results presented in Table 6 and the statistical analyses reported in the Appendix confirm that—even at the more difficult K/R and R/N junctures, subjects are able to perform better than chance. That is, when presented with a knowing or a reckless scenario, subjects are not just guessing, as they might do on a multiple choice exam question. If subjects were guessing, we would see something closer to 20% sorting rates because they would guess equally between the five mental states. But subjects do not exhibit this behavior. They are able to make some coarse distinctions (that is, they generally can tell that K and R scenarios are not blameless scenarios), but they run into trouble with the more precise categorization. In other words, when evaluating K and R scenarios, subjects are confused—more so than in P and B scenarios—but not clueless.

Nevertheless, it is clear that the success rate of sorting K and R may go a long way in explaining the indistinguishable punishment ratings seen in Experiments 1 through 3. It seems that subjects cannot sort K and R scenarios nearly as well as the MPC presumes they can.

TABLE 6: SORTING SUCCESS RATE IN EXPERIMENT 4 ("CAN SUBJECTS DISTINGUISH BETWEEN MENTAL STATES?"), BY MENTAL STATE

	Correct Mental State: Purposeful	*Correct Mental State:* Knowing	*Correct Mental State:* Reckless	*Correct Mental State:* Negligent	*Correct Mental State:* Blameless
Subject chose: Purposeful	78%	9%	5%	2%	0%
Subject chose: Knowing	14%	50%	42%	5%	1%
Subject chose: Reckless	5%	30%	40%	31%	3%
Subject chose: Negligent	2%	10%	12%	48%	8%
Subject chose: Blameless	1%	2%	1%	15%	88%

What To Notice in Table 6: Subjects' success rate at correctly identifying knowing or reckless scenarios is significantly lower than the success rate of identifying purposeful and blameless scenarios, respectively. Negligent scenarios are also more difficult for subjects to identify than purposeful or blameless scenarios.

Note: The shaded cells in Table 6 display the sorting success rate for each mental state. The non-shaded cells display the percentage of subjects across the other four (incorrect) options. For instance, looking at the column labeled "Purposeful," 78% of subjects correctly identified these scenarios; 14% mistook them for knowing; 5% mistook them for reckless; 2% mistook them for negligent; and 1% mistook them for blameless.

4. *Results from Experiment 5*: *"How Do Subjects Punish After They Have Practiced Sorting Mental States?"*

Having established in Experiment 4 that sorting ability was significantly worse for the K and R scenarios, we turned in Experiment 5 to a combination of sorting and rating tasks. This experiment explored whether the task of punishment rating might be interacting with the

task of sorting—that is, do subjects become better raters when they are first asked to sort? To answer this question, subjects in Experiment 5 were first instructed to sort fifteen questions according to the MPC definitions (these fifteen questions were identical in form to those used in Experiment 4). After the sorting questions, subjects were given fifteen punishment rating questions with different scenarios than the ones they sorted (these rating questions were presented without MPC definitions, as they were in Experiment 1).[91] Experiment 5 thus allowed us to determine whether a subject's punishment ratings were related to the subject's sorting ability. We were able to test whether "good sorters" punished differently than did "bad sorters."

The results from Experiment 5 suggest that the inability to distinguish K from R in punishment ratings most likely stems from an inability to distinguish K from R in sorting. Once again, the ratings for knowing and reckless cluster together, sitting lower than purposeful (see Figure 3 for the K vs. R comparison). Negligent ratings are below the K/R cluster, and blameless is at the bottom. But what about the "good" sorters? When we run the analysis again, but limit our scope to just those sorters who are correct 75% of the time or more, we still do not find a significant difference between knowing and reckless. When we restrict our analysis to the "bad sorters" (those correct less than 50% of the time) the K/R distinction is again blurred. Thus, while it may be that those who understand and can utilize the definitions (that is, those who can sort correctly) punish slightly differently for K than for R, even these good sorters fail to make crisp distinctions between the two.

IV
IMPLICATIONS

What do the results of our experiments suggest about the utility of the MPC culpability categories, and the need to reform those categories or the way they are conveyed to jurors? At the outset, we must suggest caution. This is but one set of experiments in a young—indeed almost non-existent—empirical literature. Future studies may point in different and unanticipated directions.

---

[91] We counter-balanced our themes for this experiment. We randomly assigned our thirty themes either into group "A" or group "B." Subjects were randomly assigned to either "sort using group A themes, then rate using group B themes" or "sort using group B themes, then rate using group A themes." As with the other experiments, question order was randomized.

A. Study Limitations

Like all experiments, ours have their limitations. Among them are the following five.

First, if it is true that average people have difficulty seeing the differences between knowing and reckless mental states, then we, as the drafters of these scenarios, may also have had that difficulty. Perhaps, given the right scenarios, our subjects could perceive a difference between knowing and reckless states of mind, but we did not construct such readily accessible scenarios. The successful results of our external validation with criminal law professors diminishes this likelihood. But criminal law professors are admittedly trained to perceive these differences.

Second, we intentionally limited the focus of our experiments to *results* elements of crimes. So it must be noted that our experiments do not speak to whether people can distinguish between when a wrongdoer "knows" versus "should have known" about *circumstance* elements of a crime, i.e., elements having to do with the existence of a particular existing or historical fact.[92] Indeed, a natural extension of the present study would be to test the MPC assumptions as they operate for circumstance elements.

Third, although our number of subjects was large, our results still depended on a relatively small number of scenarios. Those scenarios raise their own ecological concerns. We tried to force our subjects to conclusions about mental states by using language specifically designed to signal those mental states. But of course this is not the kind of evidence real jurors see in real cases. Witnesses never testify that a defendant was "consciously aware of a substantial risk." Real jurors must use much more mundane evidence—largely in the form of the act itself and the circumstances surrounding the act—to infer mental states. Having said this, it should be noted that if our signaling language artificially over-led subjects to the desired category, that would make our findings at the K/R boundary even *more* troublesome. And it would make the ability of subjects to distinguish all the *other* categories correspondingly less impressive.

Fourth, our specific implementations of the MPC language may be imperfect. For instance, because the MPC describes reckless action as involving "conscious disregard" of a risk, we generated reckless scenarios that involved both an awareness of a risk and a *choice* to go ahead and take the risk. Such language of choice appears in many of our R scenarios, but not in the K scenarios. Subjects may have viewed

[92] *See supra* note 17 (discussing distinction between circumstances and results elements).

this conscious choice more negatively than knowledge of certain harm, and punished accordingly. Future studies can be designed to explore this possibility.

Of course, part of the challenge in developing mental state signals suitable for experiments is the fact that the MPC is not clear on how its probabilistic language (such as "high probability") should be interpreted. For example, does the phrase "practical certainty" in the definition of "knowing" mean "close to 100%," or "90%," or something else? Would eight chances in ten be sufficient?[93] While our rotation of signaling language ensures that no particular phrase drives the overall results, it remains the case that ambiguities—both in the MPC language itself and in the specific mental state signals we employed—may partially explain some of the confusion around the K/R boundary. For instance, among our signals for the undefined term "substantial risk" in the recklessness mental state definition are the phrases "very likely" and "could easily happen." Reasonable people could differ about whether the MPC's use of "substantial" means "more likely than not," as our signaling language could be read to imply.[94]

Fifth, there is the ever-present and important caution that the descriptive is never by itself prescriptive.[95] So even if it turns out to be true that humans can distinguish P, K/R, N, and blameless states of

---

[93] Kenneth W. Simons, *Should the Model Penal Code's* Mens Rea *Provisions Be Amended?*, 1 OHIO ST. J. CRIM. L. 179, 183 (2003) (raising these questions and arguing that "[e]ven if we are uncomfortable employing a precise number, greater clarity would be valuable").

[94] It is not the purpose of this Article to explore the many reasons why subjects might have difficulty at this K/R boundary. But there are several possibilities. First, there are categorical distinctions between purpose and non-purpose (the P/K boundary) as well as between awareness and non-awareness (the R/N boundary), but no such category between degrees of awareness (the K/R boundary) where the distinction is one of degree on a spectrum rather than a category. Second, our subjects may punish in part by imagining how fearful they would be to have our hypothetical John running around in their community, and a reckless actor may be more threatening. Third, "knowing" sounds like something good, while "disregarding" sounds like something bad.

[95] This is what moral philosophers call "the naturalistic fallacy," or sometimes "Hume's gap" after the Scottish philosopher David Hume. For an analysis of how the science of moral realism might help legal theorists cross Hume's gap, see Morris B. Hoffman, *Evolutionary Jurisprudence: The End of the Naturalistic Fallacy and the Beginning of Natural Reform?*, in 13 CURRENT LEGAL ISSUES: LAW AND NEUROSCIENCE 483 (Michael Freeman ed., 2011). Scholars differ on whether this kind of moral realism might inform our punishment theories and practices. *Compare* Donald Braman, Dan M. Kahan & David A. Hoffman, *Some Realism About Punishment Naturalism*, 77 U. CHI. L. REV. 1531, 1532 (2010) (criticizing punishment naturalism by arguing that although moral judgments depend on numerous cognitive and physiological mechanisms that are presumably a product of evolutionary pressures, they are not innate insofar as they depend crucially on social meaning that varies across cultural groups), *with* Paul H. Robinson, Owen D. Jones & Robert Kurzban, *Realism, Punishment, and Reform*, 77 U. CHI. L. REV. 1611, 1613 (2010) (responding that whatever the source of people's shared intuitions of justice, those

mind from one another, and assign punishment to those different states as the MPC suggests, that does not necessarily answer the policy question of whether these "natural" categories should continue to be given legal traction. Indeed, strict liability crimes depart from mental state categories altogether, natural or otherwise. Strict liability crimes, such as drunk driving or the increasingly long list of regulatory crimes, usually reflect policy decisions.

Likewise, even if it turns out that humans are very poor at detecting differences between K and R, that does not necessarily mean that legislatures should throw these categories out. The categories may still serve important policy functions, some of which are discussed below. Either way, our results suggest that if this distinction is to continue to matter, legislatures and courts will have to do a better job of articulating it in their codes and jury instructions. But in deciding all these policy questions, and subject to the caveats already mentioned, surely we cannot ignore these descriptive results, even if the results themselves are insufficient to drive any particular prescriptive conclusions.

B. Lessons About Culpability

What policy lessons might be drawn from our results? First, the (somewhat) good news for the MPC is that it seems the empirical cup is more than half full. Just how good this news is depends, of course, on one's expectations of and aspirations for the criminal justice system. On the one hand, subjects performed substantially better than chance at recognizing and appropriately punishing all levels of culpability, except for K/R, even without jury instructions. This suggests that the MPC has done a reasonably good job at reflecting our intuitions about blameworthiness. And it should go a long way toward answering those who level the broadest criticism at the MPC approach in continuing to claim that the MPC categories are over determined if not wholly fictitious.[96]

On the other hand, a system that expects jurors to be able to tell the difference between purposeful, negligent and blameless acts, that requires prosecutors to prove those states of mind beyond a reasonable doubt, and that sometimes attaches great significance to these differences, should not be too complacent about our results. Although

shared intuitions are something to which system designers and social reformers would be wise to give special attention).

[96] *See supra* note 37 and accompanying text (listing critics who have argued that one or more of the MPC culpability categories overlap).

our subjects performed well above chance, they were still, for example, able to recognize negligent behaviors only 48% of the time.

Our other important finding was that subjects were less able to correctly identify K and R scenarios than any of the other mental states, even when told how to do so (by receiving the definitions of K and R). And even when they could distinguish K from R, they did not punish any differently. These findings, if validated in future studies, will demand attention and perhaps reform.

Although state criminal codes do not contain a large number of crimes for which the distinction between K and R matters, there is one kind of crime—homicide—where the distinction is critical for sentencing outcomes.[97] In many MPC states, the sentencing differences between a knowing homicide and a reckless one are enormous.[98] In Colorado, for example, an MPC state in which the judge co-author of this paper presides, a knowing murder is called second degree murder and carries a mandatory prison sentence of between sixteen and forty-eight years.[99] A reckless murder, by contrast, is called manslaughter

---

[97] Other important MPC crimes that require knowing conduct, and which are not proved by mere recklessness, include felonious restraint (§ 212.2), false imprisonment (§ 212.3), sexual assault (§ 213.4), and false reporting (§ 241.5). MODEL PENAL CODE §§ 212.3, 213.4(1)–(3), 220.1 (1962). Perhaps the relative infrequency of K/R crimes (that is, crimes that will trigger one level of seriousness if committed knowingly and a lesser level of seriousness if committed only recklessly) is itself an implicit recognition that people have difficulty with these two categories.

[98] Most 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. *See, e.g.*, ARIZ. REV. STAT. ANN. §§ 13-1103(A)(1), -1104(A)(2) (2011) (comparing second degree murder to manslaughter); ARK. CODE ANN. §§ 5-10-103(a)(1), -104(a)(3) (2011) (same); HAW. REV. STAT. ANN. §§ 707-701.5(1), -702(1)(a) (LexisNexis 2011) (same); 720 ILL. COMP. STAT. ANN. 5/9-1(a)(1)–(2), 5/9-3(a) (West 2011) (comparing first degree murder to manslaughter); IND. CODE ANN. §§ 35-42-1-1(1), -5 (West 2011) (comparing murder to reckless manslaughter); ME. REV. STAT. ANN. tit. 17, §§ 201(1)(A), 203(1)(A) (2010) (comparing murder to manslaughter); MO. REV. STAT. §§ 565.021(1)(A), -.024(1)(1) (2011) (comparing second degree murder to involuntary manslaughter); N.H. REV. STAT. ANN. §§ 630:1-b(I)(a), :2(I)(b) (2011) (same); N.J. STAT. ANN. §§ 2C:11-3(a)(2), :11-4(a)(1) (West 2011) (comparing first degree murder to aggravated manslaughter); TENN. CODE ANN. §§ 39-13-210(a)(1), -215(a) (2011) (comparing second degree murder to reckless homicide); TEX. PENAL CODE ANN. §§ 19.02(b)(1), 19.04(a) (West 2011) (comparing murder to manslaughter); UTAH CODE ANN. §§ 76-5-203(2)(a), -205(1)(a) (LexisNexis 2010) (same).

[99] Second degree murder, without any heat of passion mitigator, is defined and classified as a Class 2 felony at COLO. REV. STAT. §§ 18-3-103(1), -103(3)(a) (2010). Class 2 felonies ordinarily carry a non-mandatory presumptive sentence of eight to twenty-four years. *Id.* § 18-1.3-401(1)(a)(V)(A). But murder can also be a crime of violence, a determination that has the effects of (1) increasing the range to sixteen to forty-eight years; and (2) making a prison sentence mandatory. *Id.* § 18-1.3-406 (pertaining to murders involving deadly weapons or to crimes causing serious bodily harm or death).

and carries a non-mandatory sentence of two to six years.[100] In the very worst case, therefore, the difference between a jury finding a knowing homicide and a reckless one is the difference between a forty-eight-year prison sentence and probation. The very *smallest* this difference could ever be is ten years—the difference between the minimum of sixteen years for a knowing homicide and the maximum of six years for a reckless one.

True, the sharpest edges of this difficulty may be smoothed somewhat by the doctrine of extreme indifference, which makes even a reckless homicide, if it was reckless enough, the equivalent of first degree murder.[101] But in the judge-author's experience, juries are rarely instructed on extreme indifference in homicide cases; the recklessness issue arises much more commonly in cases that start out as second degree murder, with the jury instructed on the lesser-included offense of reckless manslaughter.

There are even special categories of homicide where the difference between knowing and reckless takes on particular significance. Child abuse resulting in death is one example. In several states, the stem crime requires only a reckless killing, but acting knowingly can make the crime substantially more serious.[102] In Colorado, for example, a reckless child abuse resulting in death is punishable by a mandatory sentence of between sixteen and forty-eight years in the penitentiary.[103] But if the prosecution can prove that the death was knowing, that the defendant was in a position of trust, and that the child was less than twelve years old, it is first degree murder punishable by a mandatory life sentence without the possibility of parole, or

---

[100] Manslaughter is defined and classified as a Class 4 felony in Colorado. *Id.* § 18-3-104. Class 4 felonies carry a non-mandatory presumptive sentence of between two and six years. *Id.* § 18-1.3-401(1)(a)(V)(A). Manslaughter is not defined as a crime of violence under § 18-1.3-406.

[101] *See supra* note 18 (explaining that most codes recognize a special kind of homicide that is so gross that the act will be treated as if it were purposeful).

[102] In Arkansas, knowingly causing the death of a person fourteen years of age or younger is a class Y felony, punishable by ten to forty years in prison. ARK. CODE ANN. §§ 5-10-102(a)(2), -102(c), -4-401(a)(1) (2011). Manslaughter, reckless killing, is only a class C felony, punishable by three to ten years in prison. §§ 5-10-104(a)(3), -4-401(a)(4) (2011). In Utah, knowing homicide incident to child abuse constitutes a capital felony punishable by death, twenty-five years to life in prison or life in prison without parole. UTAH CODE ANN. §§ 76-3-206(1), -5-202(1)(d), -5-202(3)(a). Reckless homicide incident to child abuse is only a first degree felony, punishable by five years to life in prison. *Id.* §§ 76-3-203(1), -5-208(1)(a), -5-208(2) (LexisNexis 2010).

[103] Knowing or reckless child abuse resulting in death (without the position of trust and age aggravators) is a Class 2 felony, COLO. REV. STAT. § 18-6-401(7)(a)(I) (2010), with a specifically aggravated sentencing range, *id.* § 18-1.3-401(8)(d).

theoretically even death.[104] Some states also make distinctions between K and R when it comes to killing police officers, firefighters, or judges. For example, in New Hampshire, the reckless killing of a law enforcement or judicial officer, like the reckless killing of any ordinary citizen, is manslaughter, punishable by a prison term of not more than thirty years.[105] But the knowing killing of a law enforcement or judicial officer is capital murder, punishable by either a mandatory life term without parole, or possibly even death.[106]

Because of the law of lesser-included offenses, these are not just hypothetical differences. In almost all states, a defendant is entitled to demand that the jury be instructed on all so-called "lesser-included offenses."[107] In most states, the prosecution has the same right.[108] This means that in virtually every case where a knowing homicide is charged, and even sometimes when a purposeful homicide is charged, the jury will also be asked whether the homicide was merely reckless.[109]

It would therefore be quite troubling indeed if these differing sentences are the product of jurors being required to make distinctions that they simply cannot make. In our experiments, we included two themes in which a victim was killed by our hypothetical

---

[104] Knowing child abuse resulting in death with the position of trust and age aggravators is a Class 1 felony. *Id.* §§ 18-3-102(1)(f), -6-401(7)(c). Class 1 felonies are punishable by life without parole or death. *Id.* § 18-1.3-401(1)(a)(V)(A).

[105] N.H. REV. STAT. ANN. § 630:2 (LexisNexis 2011).

[106] *Id.* § 630:1.

[107] Michael H. Hoffheimer, *The Future of Constitutionally Required Lesser Included Offenses*, 67 U. PITT. L. REV. 585, 638 (2006) (noting that the standard in federal and most state courts is to allow lesser included offense instructions in "cases where there is not only evidence sufficient to support a conviction of the lesser included offense but where there is also a real dispute about the element that differentiates the greater and lesser included offenses"). States differ on how they define a "lesser included" offense. There are two principal tests. Some states use the "elemental" or "statutory" approach: Crime *Y* is a lesser-included of Crime *X* if all the elements of Crime *Y* are also elements of Crime *X*. Thus, simple robbery (the knowing taking of a thing of value using force or threats of force) is a lesser-included offense of aggravated robbery (the knowing taking of a thing of value using force or threats of force by way of a deadly weapon). Most states, however, use the "evidentiary" or "cognate" test for lesser-includedness: Under the facts, could reasonable jurors convict the defendant of the lesser and acquit him of the greater offense? *See* Christen R. Blair, *Constitutional Limitations on the Lesser Included Offense Doctrine*, 21 AM. CRIM. L. REV. 445, 447–51 (1984) (outlining doctrine).

[108] 3 CHARLES ALAN WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 515 (4th ed. 2011).

[109] Of course, the bulk of criminal cases are plea bargained. But plea bargaining happens, as Justice Breyer so cogently put it, in the shadows of the trial. United States v. Booker, 543 U.S. 220, 255 (2005). And the shadows of trial contain the looming omnipresence of culpability.

protagonist John.[110] Looking at the K and R sorting in these two themes, subjects were correct only about 50% of the time. What we might do about the K/R problem depends on the nature of the confusion, which could have several explanations.

It may be that people do not view K primarily as a desire-based wrong but rather see it as a risk-taking wrong. Under this view, what is wrong about my shooting at the bird by aiming through you is not that I am willing to cause your death as a side effect of my desire to kill the bird, but rather that I am willing to take a big risk that you will die when I shoot the bird. Indeed, the MPC defines knowing as a risk-based state of mind and not, as the philosophers might denote it, as a side effect. This explanation not only accounts for why subjects could not distinguish K from R (because both are about risk-taking, and subjects simply see no difference between an "almost certain" risk and a "substantial" risk), it also nicely accounts for why they are so robustly able to distinguish between P and K (because the former is a desire-based wrong and the latter, in this account, a risk-based wrong).

Alternatively, perhaps subjects view R as more of a desire-based wrong than a risk-taking wrong: If we do an act conscious of a substantial risk of harm, then we desired the harm. This alternative would explain why subjects did reasonably well at distinguishing R from N (because the latter is a risk-taking wrong, and the former, in this account, is a desire-based wrong). But we are not really in dire need of an alternative explanation for the R/N junction because the difference seems palpable: Reckless actors are conscious of the risk they are taking and negligent ones are not. Moreover, closing the K/R confound in this direction would run counter to the behavioral and philosophical literature on side effects, which recognizes a clear moral distinction between intending harm and being willing to cause harm as a side effect of some other intention.[111]

Conceptually, it might also be the case that subjects are making differentiations based on their assessments of the actor's (implied) willfulness in being ignorant. The MPC's treatment of willful blindness has been debated for many years by commentators.[112] At issue is

---

[110] In one, a victim died in a car crash due to faulty brakes installed (knowingly or recklessly) by John. In the other, two skiers were killed by an avalanche started (knowingly or recklessly) by John.

[111] For a review of philosophical and experimental literature, see *supra* note 15 and accompanying text.

[112] *See, e.g.*, Jonathan L. Marcus, *Model Penal Code Section 2.02(7) and Willful Blindness*, 102 YALE L.J. 2231, 2231 (1993) (providing summary of how "[c]ourts and criminal law scholars have struggled for decades to sort out the relationship between the basic concept of knowledge, which is central to our notions of criminal responsibility, and the

whether and when the law should treat actors without "knowledge" as if they in fact had knowledge.[113] Given that such vociferous debate exists within scholarly circles about where the boundary line exists for knowing acts, it may not come as a surprise that our subjects have more difficulty assigning punishment in this area.

Better instructions might help create some discernable separation between K and R. For example, a better K instruction might emphasize the side-effect nature of K. Instead of defining K as being "aware" that an act will "almost certainly" cause the harm, this kind of improved instruction might instead define K as "not desiring the harm, but being willing to cause it in order to accomplish some other purpose." Care, of course, needs to be taken that by tinkering with K in this fashion we do not create a confound between K and P. After all, a mafia hit man may not exactly "desire" the target's death—his "desire" is to get paid, not for the target to die. This should not make the death a less culpable "side effect," because the hit man's purpose is still to kill.

Likewise, a better R instruction might be crafted to lessen the risk from "substantial" to something like "palpable" or "evident." Here again, gaining more definitional separation between K and R in this fashion may risk less separation between R and N, though these two states of mind seem safely separated by R's requirement of being conscious of the risk.

If none of these less drastic solutions help, perhaps we should consider abolishing the distinction between K and R, at least in the many states where that distinction takes on central significance in determining the degrees of homicide. If jurors cannot really tell these two categories apart, then at worst we are subjecting some similarly situated homicide defendants to divergent consequences based on the

---

concept of 'willful blindness'"); *see also* Robin Charlow, *Wilful Ignorance and Criminal Culpability*, 70 TEX. L. REV. 1351, 1372 (1992) ("No single definition of knowledge is universally agreed upon or regularly employed, even within the limited context of criminal mens rea."); David Luban, *Contrived Ignorance*, 87 GEO. L.J. 957, 962 (1999) (arguing that the MPC's emphasis on an actor's subjective state of mind at the moment of the crime is a completely different issue from that of determining willful ignorance).

[113] *See* Steven P. Garvey, *What's Wrong with Involuntary Manslaughter?*, 85 TEX. L. REV. 333, 371–72 (2006) ("An otherwise reckless actor who could have easily gathered the additional information needed to transform his belief that $p^*_{\text{substantial}}$ into the belief that $p^*_{\text{practically certain}}$ (and thus into knowledge) is willfully ignorant if he chose not to do so because he wanted for no good reason not to know."). Moreover, it is also the case that in some contexts, such as securities fraud, reckless conduct can actually be used as proof of knowledge. *Cf.* William H. Kuehnle, *Secondary Liability Under the Federal Securities Laws—Aiding and Abetting Conspiracy, Controlling Person, and Agency: Common-Law Principles and the Statutory Scheme*, 14 J. CORP. L. 313, 328 n.75 (1988) (noting that some courts appear to limit recklessness to the role of evidence of knowledge rather than as an objective standard of liability).

vagaries of a meaningless distinction. At best, we are inviting jurors to compromise their verdicts: A false K/R choice may allow jurors to exchange a vote of "not guilty" for a vote of "guilty but only of the lesser reckless charge." In any event, the legitimacy of the law is at risk when it makes such serious consequences depend on a determination that seems to have such little cognitive traction with its citizens. Thus, as the American Law Institute revisits sentencing provisions of the Model Penal Code, perhaps it should consider unwarranted sentencing disparities emerging from confusion at the K/R boundary.[114]

On the other hand, there are several arguments in favor of retaining the K/R difference even in the face of our results here, and even if these results cannot be avoided by better definitions or instructions. First, compromise verdicts are not necessarily a bad thing. Perhaps the criminal law is wise to retain the K and R categories so that jurors, and in fact even charging prosecutors, have more options.

Second, keeping R as a separate level of culpability may also insulate us from one of the deepest, and most long-running, debates in the criminal law—the extent to which the state should criminalize merely negligent behavior. R gives us a way to limit the state's apparently insatiable desire to criminalize bad judgment to circumstances of *really* bad judgment. Especially when both recklessness and negligence go to the jury, perhaps knowing about the higher category of recklessness will increase the jury's willingness to nullify on the negligence charge: "We know the judge told us the prosecution only has to prove negligence, but there is this whole other category of crime where a defendant consciously disregards a known risk. We are willing to criminalize *that*, but not mere negligence."

Third, and finally, the debate need not be limited only to keeping or jettisoning K and R as separate *general* categories of culpability. Legislatures may well have perfectly good reasons to decide that a particular kind of "knowing" crime is more blameworthy than a particular kind of "reckless" crime. But our results suggest that if this distinction is to continue to matter to legislatures, then either legislators in their code definitions, or courts in their instructions, will have to do a better job of articulating it. Our results suggest that the current options for jury instructions, driven by MPC language, are not likely to have an effect on improving the ability of jurors to correctly judge the K/R distinction. New types of instructions may prove more effective.

---

[114] On the American Law Institute's revision of MPC sentencing guidelines, see generally MODEL PENAL CODE: SENTENCING (Tentative Draft No. 1 2007) and Kevin R. Reitz, *American Law Institute Model Penal Code: Sentencing Plan for Revision*, 6 BUFF. CRIM. L. REV. 525 (2003).

Conclusion

At fifty years old, the Model Penal Code has managed to avoid rigorous empirical evaluation of two fundamental assumptions that underlie its culpability architecture. Can typical jurors, either on their own or at least when instructed, accurately sort culpable mental states into the four MPC categories? If so, are people's punishments consistent with the hierarchy of severity assumed by the MPC?

Our experiments suggest that the answers to both of these questions are a qualified "yes" for most, but not all, of the MPC states of mind. Subjects punished across the purposeful, negligent, and blameless categories in the way the MPC hierarchy assumes—purposeful conduct was punished more than negligent conduct, and negligent conduct more than blameless conduct. Not only did these blaming patterns persist across harm levels, but punishment levels also increased with harm.

In contrast to these generally confirming results, however, subjects performed much more poorly at the knowing/reckless juncture. Subjects' punishment patterns were a far cry from the MPC's expectations. For the vast majority of the themes, there was no statistically significant difference between knowing and reckless punishment ratings. In many instances, subjects actually reversed the MPC hierarchy and punished reckless behaviors *more* than they punished knowing ones. Part of this failure to differentiate punishment, we found, is likely related to subjects' inability to identify knowing and reckless scenarios as well as they identify purposeful and blameless scenarios. Subjects were able to identify purposeful conduct an impressive 78% of the time, and blameless conduct an even more impressive 88% of the time.[115] They were less good, however, at identifying knowing (50%) and reckless (40%) conduct. These poor results emerged *even when subjects were repeatedly instructed on the distinction*.[116] While subjects were not just randomly guessing (in which case they would have been accurate just 20% of the time), and while we cannot expect 100% accuracy, surely these surprisingly low rates of accuracy should give us pause.

If the knowing/reckless findings reported here are confirmed in subsequent studies, and if we as a society value treating similarly situated defendants alike (or at least non-arbitrarily), then we need to do a better job of defining these two categories. If better definitions do not solve the problem, we should seriously consider abandoning the distinction between knowing and reckless conduct, at least in cases,

115 *See supra* Table 6.

116 *Id.*

such as homicide, where that supposed distinction continues to have enormous legal significance.

Appendix A: Technical Details

This Appendix provides additional detail on the research design employed in our study, the statistical procedures used to analyze the data, and the results of the statistical analyses.

Subjects' compliance with task instructions is of special concern with online experiments because subjects cannot be monitored while engaged in the experimental tasks. To address this issue, experimental psychologists have developed "attention filters" designed to ascertain whether subjects are in fact following instructions and paying attention to the material being presented to them online. In our experiments, we employed a modified version of the filter developed by psychologist Daniel Oppenheimer and his colleagues.[117] The design of the attention filter question was such that users who did not read carefully would see, in large font, a headline reading "Background Questions on Sources for News" as well as another large, bold question: "From which of these sources have you received information in the past month?" A series of check-box options were provided (for example, local newspaper, local TV news). Subjects reading carefully, however, were instructed *not* to check any of the boxes, but instead to type "123" into the text box provided.[118] The results presented in this Article are based only on the "good" subjects—those subjects who were paying attention.

A. Details of Confirmatory Statistical Analysis

In the body of the Article we present a series of graphical figures. Here we present statistical analysis that provides more detail than can be offered in the graphical presentations.

Looking first at punishment ratings in Experiment 1 ("How Do Subjects Punish with No MPC Instructions?"), we ran an Ordinary Least Squares (OLS) regression of standardized punishment on mental state. The regression model employs clustered (on subject)

---

[117] *See* Daniel M. Oppenheimer, Tom Meyvis & Nicolas Davidenko, *Instructional Manipulation Checks: Detecting Satisficing To Increase Statistical Power*, 45 J. Experimental Soc. Psychol. 867, 867–68 (2009) (describing filter in which subjects must carefully read instructions which, counter to boldface headline above instructions, tell subjects not to actually click on answer to question).

[118] Across the five experiments, 45% of subjects successfully answered the attention filter question. Additional analysis suggests that even when including the "bad" responses, substantive effects do not differ sharply. Moreover, a question for speculation is whether jurors are more like the "good" subjects, the "bad" subjects, or somewhere in between.

robust standard errors, to account for the fact that punishment ratings are not independent, but rather grouped by subject (i.e. subjects rated thirty scenarios each). Post-estimation tests confirm our intuitions from Figures 1 and 2. There is a statistically significant difference in standardized punishment rating between purposeful and knowing scenarios ($F(1, 195) = 165.95$, $p < .01$); between reckless and negligent ($F(1, 195) = 203.15$, $p < .01$), and between negligent and blameless ($F(1, 195) = 616.15$, $p < .01$).[119] While these P/K, R/N, and N/B differences were consistently statistically significant, the statistical significance of the difference between knowing and reckless punishment ratings was not consistent across model specifications.[120] Moreover, the substantive difference in average K and R ratings is quite small.[121] If we examine the results theme-by-theme, in the vast majority of themes no statistically significant difference emerges between K and R scenarios (see Appendix Table A1, available online at: http://law.vanderbilt.edu/download.aspx?id=6606).[122]

The variation that is evident across themes may have to do with differences in the scenario fact patterns (over and above the mental state signaling language) that make mental states more (or less) discoverable.

In Experiment 2 ("How Do Subjects Punish After Reading the MPC Definitions Once?"), as in Experiment 1, there is a statistically significant difference in mean standardized punishment rating

[119] An alternative model specification employing punishment ratings as the dependent variable and controlling for theme-specific effects also produced substantively similar results regarding these differences. In addition, expanded models were run to explore the influence of harm level, as well as additional subject-level variables, on punishment ratings. An ordinal logit model was constructed to explain punishment ratings (non-standardized). In addition to the dummy variables for purposeful, knowing, reckless, and negligent (blameless as the baseline), the model included a standardized measure of theme harm level, and measures of subjects' age, race, political ideology, education level, and past experience as a crime victim. In this expanded model, harm levels acted as expected, with greater harm levels producing higher punishment ratings.

[120] There are statistical models in which a statistically significant difference between K and R can be seen. For instance, when specifying punishment (not standardized) as the outcome variable, and introducing theme-specific dummy variables, the difference between K and R punishment ratings is significant ($?2(1) = 9.81$, $p < .01$). In the model discussed *supra* note 119, the difference is also significant ($?2(1) = 9.76$, $p < .01$).

[121] Even though statistically significant in some models, the substantive difference between the two ratings is quite small, with an average K punishment of 4.9 and an average R punishment of 4.8.

[122] Although there are some themes for which a statistically significant difference is found between K and R, the process of making multiple comparisons (that is, comparing thirty times, once for each theme) introduces a greater number of false positive findings. That is, by making thirty comparisons instead of one (pooled across all themes), it is more likely that we will find a difference between K and R. Thus, we cannot say based on our data alone that there is necessarily a difference between K and R in these themes.

between purposeful and knowing scenarios ($F(1, 95) = 109.49$, $p < .01$); between reckless and negligent ($F(1, 95) = 122.62$, $p < .01$), and between negligent and blameless ($F(1, 95) = 319.15$, $p < .01$). As in Experiment 1, no statistically significant difference consistently emerges between knowing and reckless punishment ratings. These results are summarized in Appendix Table A2, available online at http://law.vanderbilt.edu/download.aspx?id=6606.[123] Taken together, the results from Experiment 2 suggest that exposure to the MPC definitions does not solve the confusion surrounding K/R punishment gradation.

In Experiment 3 ("How Do Subjects Punish When They Have Continuous Access to the MPC Definitions?"), we made the MPC definitions available to subjects throughout their rating exercise. That is, when selecting a punishment level, subjects also saw the MPC definitions presented on the bottom of their computer screen. Under these modified conditions, we again saw that there were statistically significant differences in punishment ratings between purposeful and knowing scenarios ($F(1, 96) = 85.80$, $p < .01$); between reckless and negligent ($F(1, 96) = 99.52$, $p < .01$); and between negligent and blameless ($F(1, 96) = 253.56$, $p < .01$).

However, as was displayed in Figure 3 and reported in more detail in Appendix Table A3, available online at http://law.vanderbilt.edu/download.aspx?id=6606, there is no marked improvement in subjects' ability to differentiate punishment ratings between the K and R scenarios. If anything, the conditions used in Experiment 3 may have exacerbated the K/R punishment rating confusion.

The results of Experiment 4 ("Can Subjects Distinguish Between Mental States?") were presented in Table 6 in the text. Here we report on two additional questions of interest. First, we address the question: Given a K or an R scenario, are subjects more likely to correctly identify it as K or R, respectively? The answer is that subjects are significantly more likely to identify K scenarios as K ($F = 16.11$, $p < .01$), and significantly more likely to identify R scenarios as R ($F = 16.11$, $p < .01$). These findings are consistent with our summary statistics, which show that subjects are able to sort even these middle-category mental states with better than chance accuracy.

Although subjects are above chance in identifying K and R scenarios from each other, a second question of interest is: Do subjects do better or worse in their sorting of K and R scenarios, relative to

[123] As in Experiment 1, the statistical significance of the K/R punishment rating varies with model specification. In the OLS model, the ratings are nearly significant at the 95% confidence level ($F(1, 95) = 3.73$, $p = .06$), but the substantive difference in overall average ratings is again very small: 5.2 for recklessness and 5.4 for knowledge.

sorting of P, B, and N scenarios? To address this second question, we performed logit regression of the correct/incorrect variable on four dichotomous variables for purpose, knowledge, recklessness, and negligence. Blamelessness was omitted from the model as the baseline. The regression model employs clustered (on subject) robust standard errors, to account for the fact that the sorting correct measures are not independent, but grouped by subject.

Post-estimation chi-squared tests find that subjects were significantly less accurate in sorting K versus P scenarios ($\chi^2(1) = 146.18$, $p < .01$), and also significantly less accurate in sorting R versus P scenarios ($\chi^2(1) = 185.40$, $p < .01$). The recklessness category in particular proved most difficult. Subjects were more accurate in sorting K versus R scenarios ($\chi^2(1) = 16.11$, $p < .01$), and were more accurate too in sorting N versus R scenarios ($\chi^2(1) = 15.61$, $p < .01$).

Several additional regression models were run to explore the influence of harm level, as well as additional subject-level variables, on the likelihood of correctly sorting mental states. As discussed previously, all regression models employ clustered (on subject) robust standard errors. Added to the logit model described above were a standardized measure of theme harm level, and measures of subjects' age, race, political ideology, education level, and past experience as a crime victim. In this expanded model, the relationships reported above still hold.

Finally, we sought to test in Experiment 5 ("Sorting Plus Rating") whether engaging in the sorting exercise before rating would produce greater differentiation in punishment ratings. Consistent with Experiments 1, 2, and 3 we found a statistically significant difference in standardized punishment ratings between purposeful and knowing ($F(1, 149) = 73.47$, $p < .01$); between reckless and negligent ($F(1, 149) = 109.59$, $p < .01$); and between negligent and blameless ($F(1, 149) = 398.53$, $p < .01$) scenarios.

But we do not see improvement at the K/R boundary. As reported in Appendix Table A4, available online at http://law.vanderbilt.edu/download.aspx?id=6606, it remains the case that—even after sorting with MPC definitions—subjects are generally unable to significantly differentiate their K/R punishment ratings.

B. Summary of Blame Rating Experiments

The studies reported in the main text relied on an outcome variable that asked subjects to rate their *punishment* level. In addition, however, we ran a parallel series of rating studies that were in every way identical except that, instead of asking for a punishment rating,

we asked for a *blame* rating.[124] We denote these Experiments 1b, 2b, 3b, and 5b with a "b" to indicate that we used the blameless rating question: "On a scale from 0–9, with 0 being not at all blameworthy and 9 being extremely blameworthy, how blameworthy is John for his behavior?"

We ran this additional series of experiments to ensure that our results were not wholly dependent on the way we asked the rating question. The results from the blame rating experiments essentially mirror our punishment rating results, giving us confidence that whether it is in assigning punishment or blame, the K/R boundary proves the most difficult for subjects to navigate.

Turning to the specific results, in Experiment 1b ("How Do Subjects Blame with No MPC Instructions?"), we find that there are significant differences in blame rating between purposeful and knowing ($F(1, 193) = 59.56$, $p < .01$); between reckless and negligent ($F(1, 193) = 213.05$, $p < .01$); and between negligent and blameless scenarios ($F(1, 193) = 1149.85$, $p < .01$). There was, however, no significant difference in blame rating between knowing and reckless scenarios ($F(1, 193) = 1.88$, $p = 0.17$).

In Experiment 2b ("How Do Subjects Blame After Reading the MPC Definitions Once?") the same pattern emerged. There were significant differences in blame rating between purposeful and knowing ($F(1, 95) = 27.40$, $p < .01$); between reckless and negligent ($F(1, 95) = 81.44$, $p < .01$); and between negligent and blameless scenarios ($F(1, 95) = 371.08$, $p < .01$). There was, again, no significant difference in blame rating between knowing and reckless scenarios ($F(1, 95) = 2.34$, $p = .13$).

In Experiment 3b ("How Do Subjects Blame When They Have Continuous Access to the MPC Definitions?"), the pattern of response remained the same. There were significant differences in blame rating between purposeful and knowing ($F(1, 93) = 55.22$, $p < .01$); between reckless and negligent ($F(1, 93) = 151.98$, $p < .01$); and between negligent and blameless scenarios ($F(1, 93) = 429.99$, $p < .01$). There was, as before, no significant difference in blame rating between knowing and reckless scenarios ($F(1, 93) = .12$, $p = .73$).

Finally, in Experiment 5b ("How Do Subjects Blame After They Have Practiced Sorting Mental States?"), we find the same pattern of results as in the previous blame rating experiments. There were significant differences in blame rating between purposeful and knowing ($F(1, 151) = 84.55$, $p < .01$); between reckless and negligent ($F(1, 151) = 133.32$, $p < .01$); and between negligent and blameless scenarios (F (1,

[124] For discussion of blame experiments, see *supra* note 83.

151) = 628.85, $p < .01$). There was, as before, no significant difference in blame rating between knowing and reckless scenarios ($F (1, 151) = 1.62$, $p = .21$).

Appendix B: Full Text of Scenarios

This Appendix provides the full set of 150 scenarios used in the experiments.

Due to the length of the full set, it is provided online at this location: http://law.vanderbilt.edu/download.aspx?id=6421.[125]

[125] For replication purposes, our data may be downloaded from this location: http://dvn.iq.harvard.edu/dvn/dv/guiltyminds