Flipping the Culpability Coin:

Where the Model Penal Code Fails Defendants

Forthcoming: 80 NYU Law Review (2011)

Francis X. Shen
Vanderbilt University – School of Law

Honorable Morris B. Hoffman
Second Judicial District (Denver), State of Colorado
& University of Colorado Law School

Owen D. Jones
Vanderbilt University – School of Law
& Department of Biological Sciences

Joshua D. Green
Harvard University – Department of Psychology

Rene Marois
Vanderbilt University – Department of Psychology

FLIPPING THE CULPABILITY COIN:

WHERE THE MODEL PENAL CODE FAILS DEFENDANTS

Forthcoming: 80 NYU Law Review (2011)

Francis X. Shen, Morris B. Hoffman,

Owen D. Jones, Joshua D. Greene, and René Marois

[Version of February 23, 2011]

Do not quote or cite without permission.

Flipping the Culpability Coin: Where the Model Penal Code Fails Defendants

Francis X. Shen,[α] Morris B. Hoffman,[β]
Owen D. Jones,[χ] Joshua D. Greene,[δ] and René Marois[ε]

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 blame. We care, for example, whether the shooter intended to shoot and injure the person he killed. We therefore ask jurors to infer the

[α] Visiting Scholar, Vanderbilt University School of Law; Associate Director, John D. and Catherine T. MacArthur Foundation Law and Neuroscience Project.

[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 Law and Neuroscience Project; 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 Law and Neuroscience Project.

[δ] Assistant Professor of Psychology and Director of the Moral Cognition Laboratory, Harvard University; Member, John D. and Catherine T. MacArthur Foundation Law and Neuroscience Project.

[ε] Associate Professor of Psychology and Director of the Human Information Processing Laboratory, Vanderbilt University, Member, John D. and Catherine T. MacArthur Foundation Law and Neuroscience Project.

Acknowledgements: We received helpful comments from Al Alschuler, Sara Beale, Stephanos Bibas, Ted Blumoff, Josh Dressler, Nita Farahany, Jeff Fagan, Pat Furman, Dena Gromet, Dan Kahan, Rob Mikos, Thomas Nadelhoffer, Bill Pizzi and Chris Slobogin, as well as from participants at conferences of the MacArthur Foundation Law and Neuroscience Project. Katherine Jan 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.

mental state of a person they do not know as he acted in a way they did not see.

That seems hard enough. But we make the task more challenging by informing jurors – typically at the end of the trial – that they must decide which of four defined culpable mental states (or one blameless one) the defendant was in when he transgressed. This is because the vast bulk of the states (38 of them) either have adopted or have been heavily influenced by the Model Penal Code (MPC), which since 1962 has divided culpable mental states into: 1) purposeful; 2) knowing; 3) reckless; and 4) negligent.

This MPC taxonomy reflects several assumptions. For example, it assumes that average people either naturally do – or at least can, when instructed – sort real world mental states into these four categories with some reasonable reliability. It further assumes that the average person will rank order these four categories by punishment in the same sequence that the MPC does, that is, from *purposeful* being the most severely punished to *negligent* being the least.

We are now in Year 50 of the MPC dynasty. Given the dramatic consequences that apportioning different mental states can generate for everyone in the criminal justice system, you might think that the underlying culpability assumptions of the MPC had – at least by 50 years in – been rigorously tested. But you would be wrong. With a few notable but partial and limited exceptions, there is nearly no empirical literature on this whatsoever. This is all the more striking when we remember that the MPC is both the principal text for teaching students criminal law and the most widespread regime that criminal defendants encounter daily, as they are tried, convicted, and sentenced.

With a grant from the MacArthur Foundation, we set out to investigate these critical, yet so far untested, MPC assumptions. Assembling an interdisciplinary team of legal scholars and scientists, we designed and conducted the first comprehensive series of experiments to address the validity of the MPC culpability assumptions, on which we report here.

The bottom line is that some critical assumptions embedded in the MPC approach, or reflected in our societal faith in it, are simply wrong. Although we do not advocate abandoning the MPC

wholesale, we believe that the disjunction between some of its underlying assumptions and the reality of juror decision-making warrant consideration of possible reforms.

Part I of this article provides, for context, a brief history of culpability theories and of the Model Penal Code. Part II describes the small existing empirical literature on juror assessments of MPC mental states, highlighting both the findings and the experimental shortcomings. Part III details the designs and results of our experiments. Part IV discusses implications for reform. Technical details of the experiments, and the full set of scenarios, appear in Appendices A and B respectively.

I. CULPABILITY AND THE MODEL PENAL CODE

Accidents happen, and it seems to be a human universal that we don't generally punish true accidental acts, only culpable ones. The idea of culpability has been with us as long as we have punished each other. Primitive societies, both ancient and extant, seem universally to recognize this moral difference between accidents and non-accidents.[1] Every ancient civilization that has left a record on the issue—including 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

[1] See, e.g., EDWARD A. HOEBEL, THE LAW OF PRIMITIVE MAN 235-36 (Harvard 1954); PAUL RADIN, THE WORLD OF PRIMITIVE MAN 248-51 (Grove 1960). On the other hand, there are many examples throughout history of strict liability crimes, though these seem to be the exception rather than the rule. See 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 the damage caused by their property, including, under Roman law, for the acts of their slaves. 2 FREDERICK POLLACK & FREDRIC MAITLAND, THE HISTORY OF THE ENGLISH LAW 470-73 (2nd edition 1968). The law of products liability is the modern version of these ideas. 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 has no mind and therefore cannot have any state of mind, can be held criminally liable. Albert W. Alschuler, *Two Ways to Think about the Punishment of Corporations*, 46 AM. CRIM. L. REV. 1359 (2009).

accidents.[2] The English precept from which we get our phrase "mens rea" ("guilty mind") was "*Actus non facit reum nisi rea sit*," which means "An act is not guilty unless the mind is guilty."[3]

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

At the accident end of this pole, it seems we have always 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, going deep into both the Roman and Anglo-Saxon-German branches of the common law.[4] Non-European societies appear also to have recognized this difference. Bantu tribesmen in South Africa recognized it long before contact with Europeans, as did the Jale of New Guinea, to name just two pre-industrial societies.[5] Accidents happen, but some accidents happen because people are not as careful as they should be.

This first cut is what in modern parlance we would call 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 our modern distinction between crime and tort.[6] There are still a handful of crimes based

---

[2] Max Radin, *Intent*, Criminal, in 8 Encycl. Soc. SCI. 126 (E. Seligman & A. Johnson, eds., 1932). 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, tr., New Vision 2007).

[3] 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. Francis Bowes Sayre, *Mens Rea*, 45 HARV. L. REV. 974, 983 & n. 30 (1932).

[4] Robinson, supra note 2, at 825-30.

[5] RALPH PIDDINGTON, 1 AN INTRODUCTION TO SOCIAL ANTHORPOLOGY 345, 348 (2nd ed., Oliver & Boyd 1958) (Bantu); HORIZONS ON ANTHROPOLOGY 316 (S. Tax & L. Freeman, eds., 2nd ed., Aldine 1977) (Jale).

[6] 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

on negligence, though they are certainly the exception rather than the rule.[7]

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 various legal systems called "recklessness," "gross negligence," "willful blindness" or other words intended to convey the idea of a level of culpability higher than mere negligence but lower than desire-based intent. Its roots clearly predate the common law, and are seen in several ancient societies, including the Jews.[8] The gist of this lower level of intentionality, and higher level of negligence, was that it is wrong for a person to harm another by taking an inordinate risk – wrong enough 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, part of moral philosophy since Aquinas, that intended harm is more culpable than harmful side effects. Such behavior is arguably different from mere negligence in that the purely negligent actor has no consciousness of the risk

---

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, which were left to private revenge. See James Lindgren, *Why the Ancients May Not Have Needed a System of Criminal Law*, 76 B.U.L. REV. 29 (1996).

[7] Negligent homicide, for example. Model Penal Code § 210.4. In fact, 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 was centered on the question of when merely negligent acts should be criminalized. See generally ROY MORELAND, A RATIONALE OF CRIMINAL NEGLIGENCE (Kentucky 1944).

[8] One of the earliest Anglo-Saxon descriptions of this kind of negligence-plus, contained in the *Laws of 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).

to which he is exposing others; indeed, the heart of mere negligence is the failure to appreciate that risk. But when an actor has some appreciation of the risk of harm, and takes that risk anyway to achieve some desired result, he is behaving in a manner qualitatively differently, or so goes this argument, and deserves more blame than if he were just inattentive.

The most recent major fault line in the law of intentionality seems to be a purely American invention, dealing with the difference between desire-based intent and a new category of recklessness-plus. This new line, grounded in the degree of risk the actor is consciously undertaking, attempts to describe 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 "knowingly" by the MPC) seems to be less blameworthy than pure desire-based harms but more blameworthy than merely taking a lower risk ("recklessly" in MPC language). 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 (knowingly), at least as this new species of culpability would have it, 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 the risk is something shy of 100% and in the latter it is effectively 100%. There appear to be no express articulations of this new recklessness-plus in any legal systems until it was first suggested in the 1940s in (unexpectedly) an American treatise on federal Indian law.[9]

Other smaller fissures in the culpability continuum have suggested themselves over the centuries. Murder, 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," "provocation" and "universal malice," though technically applicable to many criminal offenses, were almost exclusively born and applied in homicide cases, with the result that they blurred even further the grades of homicide based on different states of mind. Murder even has its own category of super-intentionality, at least in the United States. In virtually every state,

[9] Robinson, supra note 1, at 846.

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."[10]

Of course, the real history of the law of culpability was considerably more confusing, and less linear, than the brief rendition in the preceding paragraphs might suggest. As governments in general, and the common law in particular, began to grope 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. But at other times, various systems tried to tease apart intentionality into the different varieties summarized above. Even when they did, different kinds of intentionality appeared at different times, and were described very 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.

If this cacophony were not bad enough, the deeply complicated question of whether the culpability inquiries are to be subjective or objective only multiplied the variations and the confusion. When we ask whether a generic defendant John was reckless, are we asking a real question about John's subjective state of mind or a group question about how all 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 was a factual inquiry not unlike what was inside a safe deposit box. 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. When we ask ourselves

[10] Interestingly, there is no second degree murder in England—the alternatives are murder for an intentional killing with or without deliberation (with the resultant life sentence) and manslaughter for everything else (which carries much less serious penalties). Parliament's Law Commission is considering adopting first and second grades of homicide American-style, distinguished by whether there was deliberation. T. Whitehead & L. Roberts, "Murderers 'to Escape' Automatic Life Sentences," *The Telegraph*, July 12, 2010.

what was in John's mind we end up asking what would have been in our minds if faced with John's situation.[11]

These are extraordinarily difficult intellectual issues in their own right, and they are only exacerbated when political bodies such as legislatures 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.[12] By 1950, this abject neglect left state criminal codes in what the United States Supreme Court famously described as "disparity and confusion [over] the definitions of the requisite but elusive mental state."[13] Commentators were less restrained, one describing state criminal

[11] 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 exhibited great confusion about whether that state of mind applied 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 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 United States 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 applied to each element, including the age of the subject, and therefore reversed Defendant's conviction. This problem is really a specific case of the larger challenge of statutory construction, and the cases, both state and federal, have generally followed the principle that if a statute has state of mind listed at the beginning of the definition, that state of mind applies to all the following elemental acts. But this continues to be an interpretive crap-shoot, often requiring courts to guess at the core nature of the crime the legislative body was trying to reach.

[12] JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 32 (Matthew Bender 2006).

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

codes as "archaic, inconsistent, unfair and unprincipled."[14] Congress did no better. It began in the early 1900s federalizing many aspects of the criminal law, and defining entirely new federal crimes, and it has never slowed down. But it has never attempted to answer these beguiling culpability issues, and indeed to this day the federal criminal code contains no general culpability definitions.[15]

It was this horribly unsettled state of the law of culpability that confronted the American Law Institute, a collection of widely-respected lawyers, judges and academics, when it began to look at criminal reform in the 1950s. Led by Harvard's Herbert Wechsler, 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 addressing: 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 still stubbornly call intentional); 2) knowing; 3) reckless; and 4) negligent. It defined them this way:

> A person acts purposefully [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.

---

[14] Sanford H. Kadish, *Fifty Years of Criminal Law: An Opinionated Review*, 87 CAL. L. REV. 943, 947 (1999). See also Herbert Wechsler, *The Challenge of a Model Penal Code*, 65 HARV. L. REV. 1097, 100-1101 (1952).

[15] 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 John S. Wiley, Jr., *Not Guilty by Reason of Blamelessness: Culpability in Federal Criminal Interpretation*, 85 VA. L. REV. 1021 (1999).

> 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].[16]

That is, the ALI retained the two oldest culpability distinctions—between 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.[17] It also set out, in its definitions of specific crimes, a general architecture that required that every crime consist of an act and one of the four levels of culpability.

It did, by the way, try to solve the subjective/objective problem: these four definitions are couched so that purposeful, knowing and reckless are subjective inquiries, and negligence an objective one.[18]

---

[16] MPC § 2.02, General Requirements of Culpability.

[17] Although it 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 murder 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. MPC § 210.3(b).

[18] See, e.g., 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). By contrast, the MPC did not attempt generally to solve the problem of how far into the elemental chain the state of mind requirement runs, except to make it clear that it runs only to "material elements" of the crime. MPC § 2.02(1). See also Paul H. Robinson &

Each specific crime definition contains a required mental state. Thus, for example, murder is defined as a purposeful or knowing killing,[19] and manslaughter as a reckless killing.[20] 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 an "almost certain" risk of harm to accomplish some other desire. Reckless is the conscious willingness to take a somewhat lower risk of harm ("substantial risk") to accomplish some other desire. And negligence is taking but being unaware of a substantial risk of harm.

Although there were many detractors,[21] the MPC formulation of culpability was hailed by most commentators as a

---

Jane A. Grall, *Element Analysis in Defining Criminal Liability: The Model Penal Code and Beyond*, 35 STAN. L. REV. 681 (1983).

[19] MPC § 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—typically requiring only a purposeful or knowing killing. That is, the MPC followed the English model in this regard. See note 10 supra.

[20] MPC § 210.3(a). Manslaughter is alternatively defined as a purposeful or knowing killing if accompanied by heat of passion. Id. at § 210.3(b). See note 15 supra.

[21] 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 Brickley, *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); MPC § 2.08(2) (negligent = reckless, at least in cases of self-induced intoxication). There have also been broader attacks, by critics who question whether the MPC missed the boat entirely by talking about a criminal's mental state as if such a mental state were a real, let alone discoverable, condition capable of doing useful legal work. See, e.g., Bruce Ledewitz, *Mr. Carroll's Mental State or What is Meant by Intent*, 38 AM. CRIM. L. REV. 71 (2001); RICHARD POSNER, THE PROBLEMS OF JURISPRUDENCE 168 (1990) ("maybe 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"). These debates are not so much about the MPC as they are a

reasonable attempt 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—38 states had largely jettisoned their criminal codes for the MPC.[22] Even in the handful of states that have not adopted it in whole or in part as legislation, it has still managed to find its way into the common law of those states because judges often turn to it for guidance. The MPC is now taught in virtually every law school, with one professor calling it "the principal text in criminal law teaching."[23] Whether in actual legislation, common law or just norms sitting inside the minds of lawyers and judges, the MPC has become, as one commentator put it, "a standard part of the furniture of the common law."[24]

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

---

continuation of the debates inside the ALI during its formulation of the MPC. In fact, some scholars have charged 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). There was, to be sure, a long period in the 1970s and 1980s during which criminal theory fell into a kind of intellectual "doldrums," as Sanford Kadish has put it. Sanford H. Kadish, *Fifty Years of Criminal Law: An Opinionated Review*, 87 CAL. L. REV. 943, 951 (1999). Whether or not the MPC was responsible for the doldrums, criminal theory is now sailing along on unprecedented gusts of interdisciplinarity. Economics, philosophy, psychology, evolutionary theory and neuroscience have stimulated a renewed interest in the foundations of criminal responsibility.

[22] Robinson & Grall, supra note 15, 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.

[23] Peter W. Low, *The Model Penal Code, the Common Law, and Mistakes of Fact: Recklessness, Negligence or Strict Liability?*, 19 RUTGERS L.J. 539, 539 (1988).

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

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 we get to those experiments, let us briefly survey the thin state of the existing empirical literature on these questions.

II. PAST STUDIES OF THE MPC MENTAL STATES

Whether jurors are capable of consistently and accurately distinguishing between the categories of mental states that the MPC identifies is an empirical question, one that legal scholarship has generally ignored.[25] What we know about jurors' ability to discern mental states must be almost entirely imported from research on the general nature of human moral reasoning, research that is not specifically tailored to the intricacies of the criminal law and that thus gives us only limited courtroom-relevant insights.

A wide body of research by social and moral psychologists, experimental philosophers, and now neuroscientists has taught us a great deal about humans' *general* ability to assess the mental states of others.[26] In particular, we have learned much about our ability to distinguish between intentional and non-intentional action, the

[25] To be sure, some commentators, such as law professor Kevin Jon Heller, have raised the question. But Heller's reflection on the state of the field is telling: "contemporary 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, 320-321 (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 (2005).

[26] Much of this research falls within a broad "theory of mind" line of research. See, e.g. WILLIAM BECHTEL, PHILOSOPHY OF MIND: AN OVERVIEW FOR COGNITIVE SCIENCE (1988). SIMON BARON-COHEN, HELEN TAGER-FLUSBERG, & DONALD COHEN, EDS., UNDERSTANDING OTHER MINDS: PERSPECTIVES FROM DEVELOPMENTAL COGNITIVE NEUROSCIENCE (2000).

basic culpability slicing that has been with us for ages.[27] For instance, there is evidence that, at least in some contexts, even infants can identify goal-motivated action.[28] But although we are naturally able to categorize some kinds of mental states,[29] the relevant question for the criminal law is more specific: are we able, either naturally or with the proper instruction, to categorize others' mental states *in the more precise manner required by the criminal law, and by the MPC in particular?*

While experimentalists like philosopher Joshua Knobe have given us insights about "people's *ordinary* criteria for intentional action,"[30] 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 historical culpability of others using the precise set of MPC definitions. On this count, psychology and philosophy findings can be translated into legally relevant

---

[27] See, e.g. BERTRAM F. MALLE, LOUIS J. MOSES AND DARE A. BALDWIN, EDS., INTENTIONS AND INTENTIONALITY: FOUNDATIONS OF SOCIAL COGNITION (2001). Bertram F. Malle & Joshua Knobe, *The Folk Concept of Intentionality*, 33 J. OF EXPERIMENTAL PSYCHOLOGY 101 (1997). Edouard Machery, *The Folk Concept of Intentional Action: Philosophical and Experimental Issues*, 23 MIND & LANGUAGE 165 (2008). Liane Young, Fiery Cushman, Marc Hauser, & Rebecca Saxe, *The Neural Basis Of The Interaction Between Theory Of Mind And Moral Judgment*, 104 PNAS 8235 (2007). F.D. Fincham & J.M. Jaspars, *Attribution Of Responsibility: From Man The Scientist To Man As Lawyer, in* ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 82 (L. Berkowitz ed., Vol 13, 1980). H.H. Kelley, *Attribution Theory In Social Psychology, in* NEBRASKA SYMPOSIUM ON MOTIVATIONS, (D. Levine ed., 1967).

[28] John H. Flavell, *Cognitive Development: Children's Knowledge About the Mind*, 50 ANNU. REV. PSYCHOL. 21 (1999). Amanda L. Woodward, *Infants' Ability To Distinguish Between Purposeful And Nonpurposeful Behaviors*, 22 INFANT BEHAVIOR & DEV. 145 (1999).

[29] Whether, and to what extent, we are born "mind readers" is debated in the fields of developmental psychology and neuroscience. Simon Baron-Cohen, *How To Build A Baby That Can Read Minds: Cognitive Mechanisms In Mindreading*, *in* THE MALADAPTED MIND: CLASSIC READINGS IN EVOLUTIONARY PSYCHOLOGY (Simon Baron-Cohen, ed. 1997).

[30] Emphasis added. Julia Kobick & Joshua Knobe, *Interpreting Intent: How Research on Fold Judgments of Intentionality Can Inform Statutory Analysis*, 75 BROOK. L. REV. 409, 420 (2009). See also: Joshua Knobe, *Intentional Action and Side Effects in Ordinary Language*, 63 ANALYSIS 190 (2003).

categories, but the translation is difficult.[31] 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."[32]

But there has been a small amount of legally relevant research into our ability to assess culpability. Nearly twenty years ago, attorney Laurence Severance teamed up with psychologists Jane Goodman and Elizabeth Loftus to conduct a small study on forty-six undergraduates at the University of Washington.[33] 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. 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."[34]

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

---

[31] B.F. Malle & S.E. Nelson, *Judging Mens Rea: The tension between folk concepts and legal concepts of intentionally*, 563 Behav. SCI. & L. *563* (2003). D. McGillis, *Attribution And The Law: Convergence Between Legal And Psychological Concepts*, 2 LAW & HUM. BEHAV. 289 (1978).

[32] 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 specific journals (e.g. Law and Human Behavior, The Behavioral Sciences and the Law, Journal of Social Psychology, Journal of experimental social psychology, etc.). Having performed this extensive search, we believe that we have identified all studies directly on point.

[33] Laurence J. Severance, Jane Goodman and Elizabeth F. Loftus, *Inferring The Criminal Mind: Toward A Bridge Between Legal Doctrine And Psychological Understanding*, 20 JOURNAL OF CRIMINAL JUSTICE 107 (1992). Surprisingly, this study has to date been cited, within the Westlaw JLR database, only once.

[34] Severance, et. al, *supra* note 33 at 108.

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." 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.[35] 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.[36] 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

[35] The definitions read: "A person acts with intent or intentionally when acting with the objective or purpose to accomplish a result which constitutes a crime. A person knows or acts knowingly or with knowledge when (1) he or she is aware of a fact, facts or circumstances or a result described by law as being a crime; or (2) he or she has information which would lead a reasonable person in the same situation to believe that facts exist which facts are described by law as being a crime. [Acting knowingly or with knowledge also is established if a person acts intentionally] A person is reckless or acts recklessly when he or she knows of and disregards a substantial risk that a wrongful act may occur and the disregard of such substantial risk is a gross deviation from what a reasonable person would exercise in the same situation. Criminal Negligence [Recklessness is also established if a person acts intentionally or knowingly] A person is criminally negligent or acts with criminal negligence when he or she fails to be aware of a substantial risk that a wrongful act may occur and the failure to be aware of that substantial risk is a gross deviation from the standard of care that a reasonable person would exercise in the same situation. [Criminal negligence is also established if a person acts intentionally or knowingly or recklessly]." Severance, et. al., *supra* note 33 at 109.

[36] An example of one of the three scenario stems used was: "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 scvera1 people." Severance, et. al., *supra* note 33 at 110.

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 (no instructions and no prompt) condition vs. (2) providing jury instructions, and vs. (3) prompting thought about one's own instructions. 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 the Washington state 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.[37] Of the four mental states examined, intent (I), knowing (K), reckless (R), and negligent (N), the only distinctions 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 reliably able to make distinctions.[38] Jury instructions made no difference in subjects' ability to make these distinctions. When assigning punishment levels, subjects were similarly able to differentiate between the extremes of intentional and negligent acts, but 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.[39]

---

[37] Severance, et. al., *supra* note 33 at 115.

[38] The researchers found that when rank ordering mental states, "legally naïve 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."

[39] In addition, Severance, Goodman, and Loftus (1992) also 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. A "template" of core elements was created for each of the four legal mental state definitions in the Washington Pattern Instructions for juries. Subject-generated definitions were then compared with each template and rated as correct or incorrect. Correct responses included all requisite elements in the template. The researchers were especially interested in subjects' pattern of 'incorrect' answers which revealed

While Severance, et. al. conducted just one study, in the early 1990s psychologist John Darley and legal scholar Paul Robinson 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.[40] In several of their studies the experimenters were interested in comparing the MPC's treatment of different culpability levels with the natural intuitions of the community. Subjects were presented with six scenarios containing instances of unconsented-to sexual intercourse, statutory rape, and property damage offenses involving damage to 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 'easy' and 'unambiguous' cases in which there were clear differences between the four variations.[41] The four variations of each scenario were given together, in reverse order of culpability level (faultlessness, negligence, recklessness, and the knowledge), and subjects then assigned liability.[42]

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 Darley and Robinson'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

---

that subjects often had their own set of preconceptions that deviated from the legal concepts of *mens rea*.

[40] PAUL H. ROBINSON & JOHN M. DARLEY, JUSTICE, LIABILITY, AND BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW (1995). Eighteen studies, designed and executed in seminars at Rutgers University School of law in Autumn of 1990 and Spring 1991, are reported in the book.

[41] 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; whereas, 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.

[42] Subjects were asked to assign liability on a scale from 0-11, with 0 corresponding to liability but no punishment and 11 corresponding to death. The liability-punishment scale also included the option of N, which corresponded to no criminal liability.

increased as the manipulated level of culpability increased.[43] The experimental results give us reason to think that, under some laboratory circumstances, individuals' mental state evaluations can be aligned with the MPC mental state hierarchy.

A decade after the Darley and Robinson study, law professor Justin Levinson conducted an experiment that explored the mediating role of culture in the assessment of defendants' mental states.[44] Levinson compared the responses of undergraduates at Beijing University in China, with those of undergraduates at UC, Berkeley and Harvard. Subjects received 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 that, for three of the four fact patterns utilized, the responses of the both the American and Chinese undergraduates did not match that predicted by the MPC.[45] Levinson also found differences between the American and Chinese responses, with the Chinese students choosing more culpable states of minds, on average, than the Americans.[46] Chinese students were also more likely to convict for attempted murder, and for assault and battery. These findings remind us of the importance of cultural variation in *mens rea* evaluations.

---

[43] Darley and Robinson also argue that there is a "special significance" in the distinction between recklessness and negligence. Subjects, they suggest, see an important distinction between "conscious" versus "inattentive" risk-taking.

[44] Justin D. Levinson, *Mentally Misguided: How State of Mind Inquiries Ignore Psychological Reality and Overlook Cultural Differences*, 49 HOW. L.J. 1 (2005). The article is a revised version of Justin D. Levinson, *Mens Rea: A Psychologically Embedded Inquiry* (unpublished L.L.M. Thesis, Harvard University) (on file with Harvard Law School Library) (2004).

[45] Although in one fact pattern distinctions were evident between purpose, knowledge, and reckless mental states, the results are not robust. When averaging over all four fact patterns, Levin finds some evidence that "participants maintained a folk mental state hierarchy," placing "purpose above knowledge above recklessness" in their punishment ratings. Levinson, supra note 25 at 20.

[46] Levinson, supra note 25 at 22.

Taken together, the studies by Severance, et. al., Darley and Robinson, and Levinson paint an incomplete, and at times contradictory, picture of the ability of jurors to evaluate criminal defendants' mental states. This might be due in part to two methodological shortcomings: (1) experimental subjects were unrealistically exposed to repeated variations of the same fact patterns and (2) the subject pools in the experiments were unrepresentative, reflecting only students and not the population more generally.

In a criminal trial, jurors are exposed to one fact pattern (albeit presented and interpreted differently by prosecution and defense). Jurors must apply the MPC's definitions in a single, particular instance. They must, for example, answer the question "Did this defendant act purposefully?" usually without the benefit of seeing evidence that the defendant did the same act, but with a different mental state.[47]

In both the Severance, et. al. and Darley and Robinson studies 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.[48] The problem with sequentially or simultaneously offering the different mental states to subjects is that it gives subjects the ability to compare. Making comparisons is a luxury individuals do not have in the jury box. Thus, we are left to wonder about questions such as: if subjects hadn't read about John acting purposefully, knowingly, and recklessly, would they have made

---

[47] In discussing his results, Levinson recognized a similar problem: "True jury trials do not present over half a dozen mental states at once. Instead, they rely on the charge and the formulation in that jurisdiction. The state of mind judgments and predictor variables I tested thus do not directly imitate the legal process. Future research should be conducted in more realistic legal situations, where fewer mental state terms are presented in any one situation." Levinson, supra note 25 at 27.

[48] Robinson & Darley concede that it would have been better methodologically to randomize the order of the variations. However, even this modification may not eliminate the bias. See Robinson & Darley, supra note 40 at 288. A better approach is to have a sufficient number of scenarios such that subjects see a certain scenario only once and make one as opposed to four mens rea judgments for each scenario.

the same judgments about him acting negligently? More fundamentally, would they even have recognized – absent the ability to compare – that this was a negligent mental state? Research designs that expose subjects to the same fact pattern multiple times cannot answer these questions. Our set of experiments improves upon these earlier research designs by exposing subjects to one and only one scenario from each fact pattern.

In addition to the multiple exposure issue, the generalizability of previous findings is limited by the nature of their samples. All three of the previous studies relied solely on student populations, either undergraduates or law students, for their subject pool. 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. Over-reliance on undergraduates has generated the term "science of the sophomore", and led to long-standing debates over the validity of studies relying solely on students.[49] To be sure, American juror pools include college students. But they are comprised primarily of individuals older than 22, and also include the 75% of Americans who do not hold a college degree.[50] In order

[49] For one critique, Steven Levitt & John A. List, *What Do Laboratory Experiments Measuring Social Preferences Reveal About The Real World?* 21 JOURNAL OF ECONOMIC PERSPECTIVES 153 (2007). For reviews of the literature, see Marc Hooghe, Dietlind Stolle, Valérie-Anne Mahéo & Sara Vissers, *Why Can't a Student Be More Like an Average Person?: Sampling and Attrition Effects in Social Science Field and Laboratory Experiments*, 628 THE ANNALS OF THE AMERICAN ACADEMY OF POLITICAL AND SOCIAL SCIENCE 85 (2010). Robert A. Peterson, *On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-Analysis*, 28 THE JOURNAL OF CONSUMER RESEARCH 450 (2001). Jerald Greenberg, *The College Sophomore as Guinea Pig: Setting the Record Straight*, 12 THE ACADEMY OF MANAGEMENT REVIEW 157 (1987). The discussion stretches back over half a century. See, e.g. Maurice L. Farber, (1952), *The College Student as Laboratory Animal* 7 AMERICAN PSYCHOLOGIST 102 (March).

[50] Based on census data from 2000, the U.S. Census Bureau reports that 24% of Americans age 25 and older had completed a college degree. U.S. Census Bureau, Educational Attainment: 2000, Report C2KBR-24 (August 2003).

to produce more generalizable findings, a more representative sample is required. Our study addresses this challenge.

III. NEW 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 protagonist. The first step in experimental design was thus to develop scenarios that were readily accessible to the subject (i.e., straightforward, reasonably believable on their face), clearly communicative of a distinct MPC mental state, and short enough so that subjects could read multiple scenarios within a reasonable time.[51] 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.[52]

Applying these principles, we drafted scenarios featuring a protagonist named John whose actions cause differing levels of harm.[53] In addition to writing specific scenarios, we also developed "themes". For purposes of this discussion, we use the term "theme," which is akin to previous researchers' "stem," to refer to the general fact pattern, e.g. John spills coffee on his neighbor's mail. We use the term "scenario" to refer to a fact pattern with a specific mental state. For example "John grabs for his pile of mail with the same hand that he's holding his coffee in, understanding that it could easily happen that some of the coffee will spill on to

[51] 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 does not come out of the blue, but his motivation is a simple, and typically neutral one. Although short, the information conveyed to subjects contains, in large measure, the central facts that are determinative of sentencing in many real world cases.

[52] For a discussion of the relationship between judgments of intentionality and harm caused, see Joshua Knobe, *The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology*, 130 PHILOSOPHICAL STUDIES 203 (2006). Edouard Machery, *The Folk Concept of Intentional Action: Philosophical and Experimental Issues*, 23 Mind & Language 165 (2008).

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

his neighbor's mail, but choosing to ignore that risk." Thus, within each theme there are five scenarios: one each for purposeful, knowing, reckless, negligent, and blameless.

We created 30 themes, ten each in one of three harm-level categories: 1) 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.[54]

Each scenario was comprised of exactly three sentences (a, b, c). Within a given theme (i.e. general fact pattern), all the "a" sentences and all the "c" sentences were identical.[55] 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 "b" sentences – which 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 counter-balanced evenly within and across all the three harm levels. The language we developed is reported in full in Table 1. We illustrate in Table 2, for one of these thirty themes, how each scenario was constructed.

[54] The full set of scenarios is available in Appendix B (available on-line at < http://law.vanderbilt.edu/download.aspx?id=6421 >).

[55] There were just a few rare exceptions where we needed to vary the word "the" to the word "a".

Table 1. 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 (achieve the specific harm)
   b. Intends (or with the intention of)
   c. Desires that
   d. Wants to
   e. Chooses to

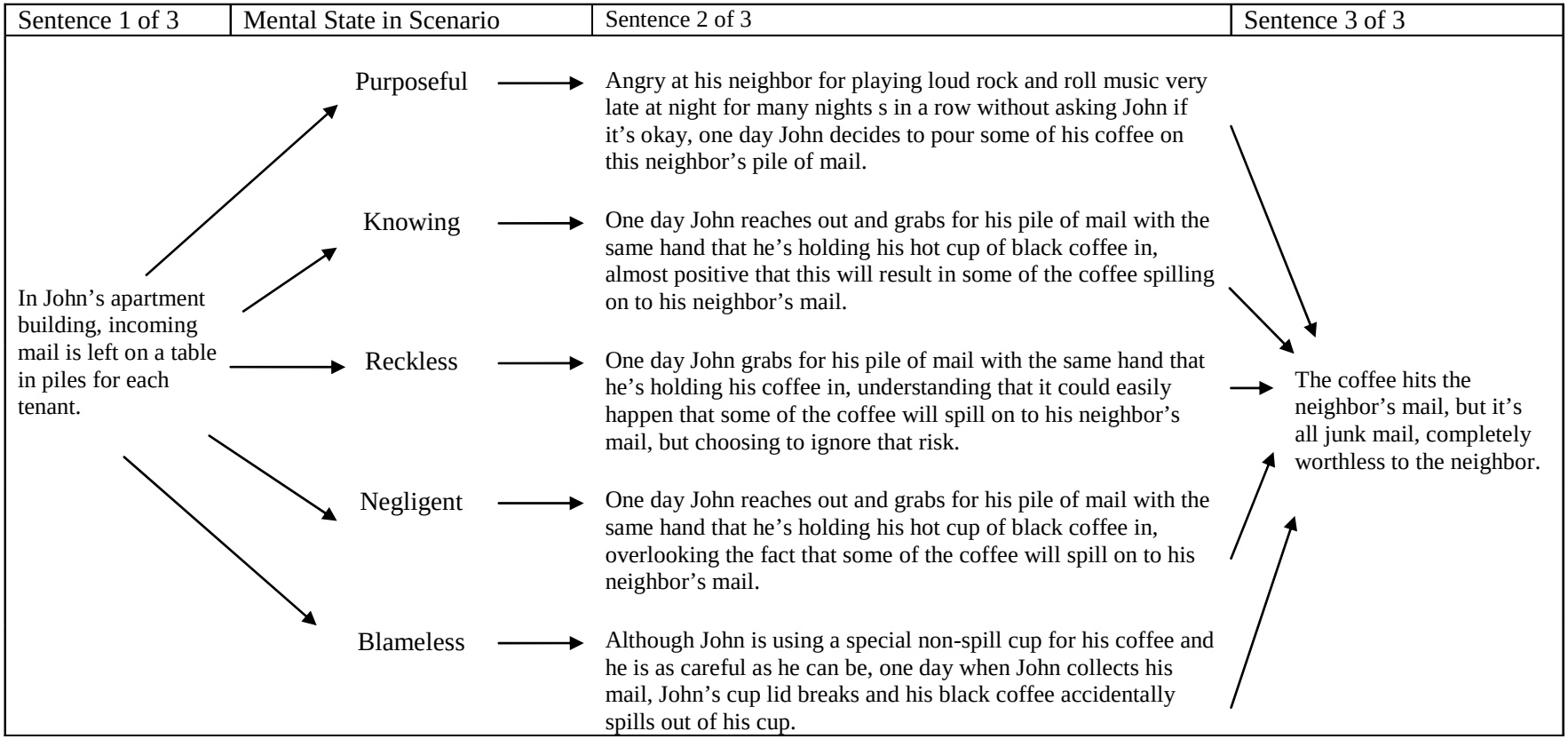
2) **Knowingly** (similar language as Purposefully, but with contextual clarification that John doesn't separately *intend* the harm that occurs; he is instead aware that acting to fulfill his separate intention *will certainly cause* (100% certain) the harm that does happen)
   a. Practically certain 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 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** (wherein harm results from:)
   a. Despite being as careful as he could, accidentally
   b. [Act is involuntary]
   c. Unavoidably
   d. Through an honest mistake
   e. Inadvertently [causes harm] despite his best efforts.

Table 2. Illustration: Varying Mental State Within A Single Theme



Note: For each of thirty themes (ten involving high harm, ten involving medium harm, and ten involving low harm), we constructed five scenarios. This is an example of one of the low harm themes. Text of all scenarios used in the experiments is presented in Appendix A. Notice that each scenario uses the same first and third sentence, varying only the middle sentence.

Signaling mental states is, of course, not just a function of word choice but also of scenario context. A particularly vexing question for scenario construction about mental states was how to structure the scenarios so that we could clearly show John's relationship to the harm being caused.[56] If we label harm as the *y* variable, and *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:

- Purposefully: John decides to do *y* via *x*
- Knowingly: John does *x*, practically certain that it will result in *y*
- Recklessly: John does *x*, aware there is a substantial risk that *y* will occur
- Negligently: John does *x* carelessly, thus causing *y*
- Blamelessly: John does *x*, and despite being as careful as he could be, he accidentally causes *y*

We turned to nine criminal law professors for external validation that our scenarios were in fact communicating the mental state we posited they did. Each professor read the 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 are indeed signaling the *mens rea* categories as defined by the MPC.[57]

---

[56] The experimental philosophy literature on intentionality reminds us that side effects (also known as the "Knobe effect" after Joshua Knobe's research) are complicated. Even if a harm happens as a side effect, as opposed to a direct effect, of a given action, if the actor knew that the side effect was going to happen, but acts anyway, we will tend to judge the actor as if he intended to cause the side effect. We return to side effects when discussing possible explanations for the conflation we find between punishment for knowing and for reckless actions. See: Joshua Knobe, *Intentional Action and Side Effects in Ordinary Language*, 63 ANALYSIS 190 (2003).

[57] It can always be argued, especially for the middle categories of Knowing, Reckless, and Negligent, it might be argued that our signals themselves are confusing. But based on our careful selection of signaling

One additional task was to validate that our assignment of low, medium, and high harm levels corresponded to subjects' perception of the harm level, so we ran a preliminary study to do that.[58] 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 the harm ratings 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 (e.g. are subjects more accurate in sorting by mental state when they are evaluating scenarios where John commits more harm?).

B. Specific Experimental Design

In order to know what added value, if any, the jury instructions giving the MPC definitions might have, we had to start with a baseline model (Experiment 1. "Punishment Without Definitions"), in which participants were given no MPC definitions to guide their rating. Participants in Experiment 1 were simply presented with a fact pattern and asked to make a punishment rating.

---

language aligned with MPC definitions, as well as the law professor external validation, we believe we have generated as clear as signals as we possibly could.

[58] 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 were randomly presented with each of the thirty blameless scenarios (i.e. the blameless version of each of the thirty different fact patterns). 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>? (E.g. How harmful is having coffee spilled on completely worthless junk mail?). Subject harm ratings were standardized within subject, to account for variance due to subject-specific factors. The standardized harm rating scores discussed in the results section 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.

We then turned to the more realistic situations of asking subjects to make punishment judgments after they heard the MPC definitions once (Experiment 2. "Punishment With One-Time Definitions"); and asking subjects to make punishment judgments while having continuous access to the definitions (Experiment 3. "Punishment With Continuous Definitions").

Building on Experiment 1, which required subjects to make ratings without the benefit of any statutory guidance on mental states, we ran two additional experiments (Experiments 2 and 3) to investigate if presenting MPC definitions would affect ratings. Table 2 presents the MPC definitions used in these two experiments.

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. As discussed above, manipulating access to the MPC definitions is consistent with variations across jurisdictions in the access jurors have to mental state definitions.[59]. This combination of experiments allowed us to see if the varying the delivery of instructions produces different patterns of punishment.

The first three experiments told us much about how individuals punish, but it left open the question of mechanisms – how had they arrived at that punishment level? In particular, we

[59] 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. Judges read the jury instructions to the jury (a common law remnant of the days when jurors were illiterate), but by culture most federal judges traditionally did not provide the jury with a written copy. The practice varies across states. Spurred in part by various jury reform efforts, the trend, even in federal courts, is not only to provide the jury with a copy of the instructions but to provide each individual juror with a copy.

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 (Experiment 4. "Sorting Mental States") to determine how well subjects were able to correctly identify each MPC category. Finally, in order to see if practice with the instructions makes for (more) perfect punishment alignment, we designed an experiment (Experiment 5, "Sorting Plus Rating") where subjects completed a mental states sorting task, and then made their punishment ratings. These five experiments, taken together, are by far the most robust test to date on the question of how MPC mental states are applied.

As we noted in Section III, one problem with previous studies is that the earlier studies exposed subjects to multiple mental states within the same fact pattern. To avoid this problem, in our experiment subjects read only 30 of the 150 short scenarios, six each from the five mental states (Purposeful, Knowing, Reckless, Negligent, and Blameless). Subjects were randomly assigned one scenario from each of thirty themes.[60] 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?"[61]

---

[60] 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.

[61] 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 JOURNAL OF APPLIED SOCIAL PSYCHOLOGY 2575 (2006). To account for this possibility, we ran a second experiment identical to the first, except for a change in the rating question asked. In Experiment 2, after reading each scenario, subjects were asked: "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 also replicated our other punishment experiments with the blame question, and in case our results were substantively the same. In the interest of space, we do not report the blame rating results.

Throughout the Article we report our findings in terms of this 0-9 scale. However, we conducted additional analysis, employed "standardized"

Table 2. 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.

---

punishment ratings, to account for the fact that individuals may interpret the 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 9 represents the death penalty, but subject B believes it represents 20 years in prison. Standardization transforms each raw punishment score (the 0-9 rating) into a "standardized" rating which can be understood as: "For scenario X, how many standard deviations, above/below the subject's mean rating (for all 30 scenarios rated) is his punishment rating?" As a practical matter the substantive results of our analysis are the same whether we use the actual punishment scores or the standardized measures. Thus, we report only the punishment scores. Results using standardized ratings are available upon request from the authors.

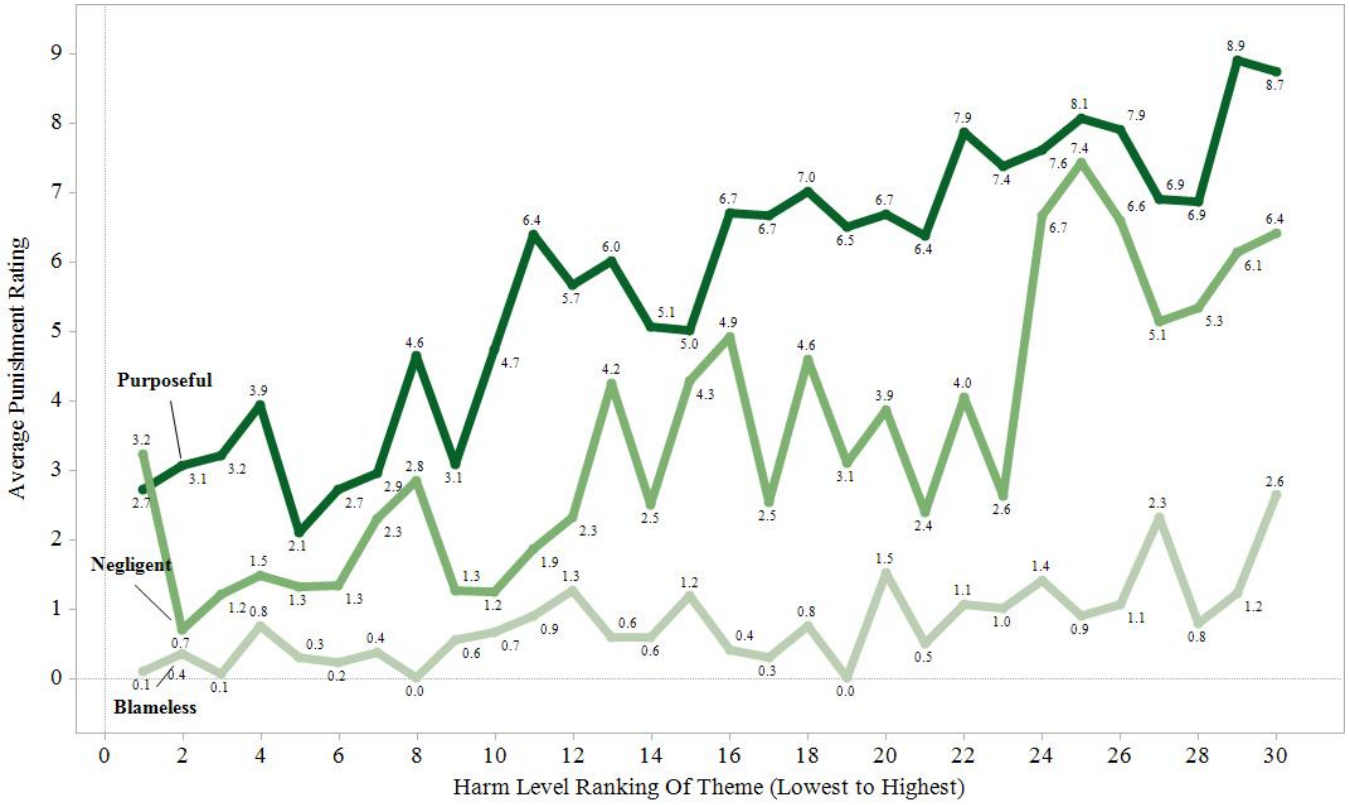
C. Results

Our results suggest that subjects punish in accordance with MPC guidelines in the purposeful, negligent, and blameless conditions (see Figure 1). Punishment ratings were highest for purposeful action. At the other end of the spectrum, blameless punishment averages were the lowest, and negligent averages were the second lowest. Where the MPC fails is at the joint of knowing and reckless (see Figure 2). The cross-cutting dark-red (K) and light-red (R) lines in Figure 2 illustrate clearly that K is often punished less harshly than R.

That we see so much back-and-forth between the two categories is powerful evidence that jurors do not see the fine grained distinctions between K and R that the MPC, and many state statutes, presume they do. The robustness of the graphical analysis was confirmed through several statistical analyses, reported in Appendix A.

Put another way, our fictional defendant John would not have been assured of systematically better treatment for the lesser crimes of acting recklessly instead of knowingly. And if John had acted knowingly, the public could not have been assured that he wouldn't have received the inaccurate, less severe punishment, for acting recklessly. This mis-alignment calls into question the MPC's effectiveness at this particular conceptual joint.

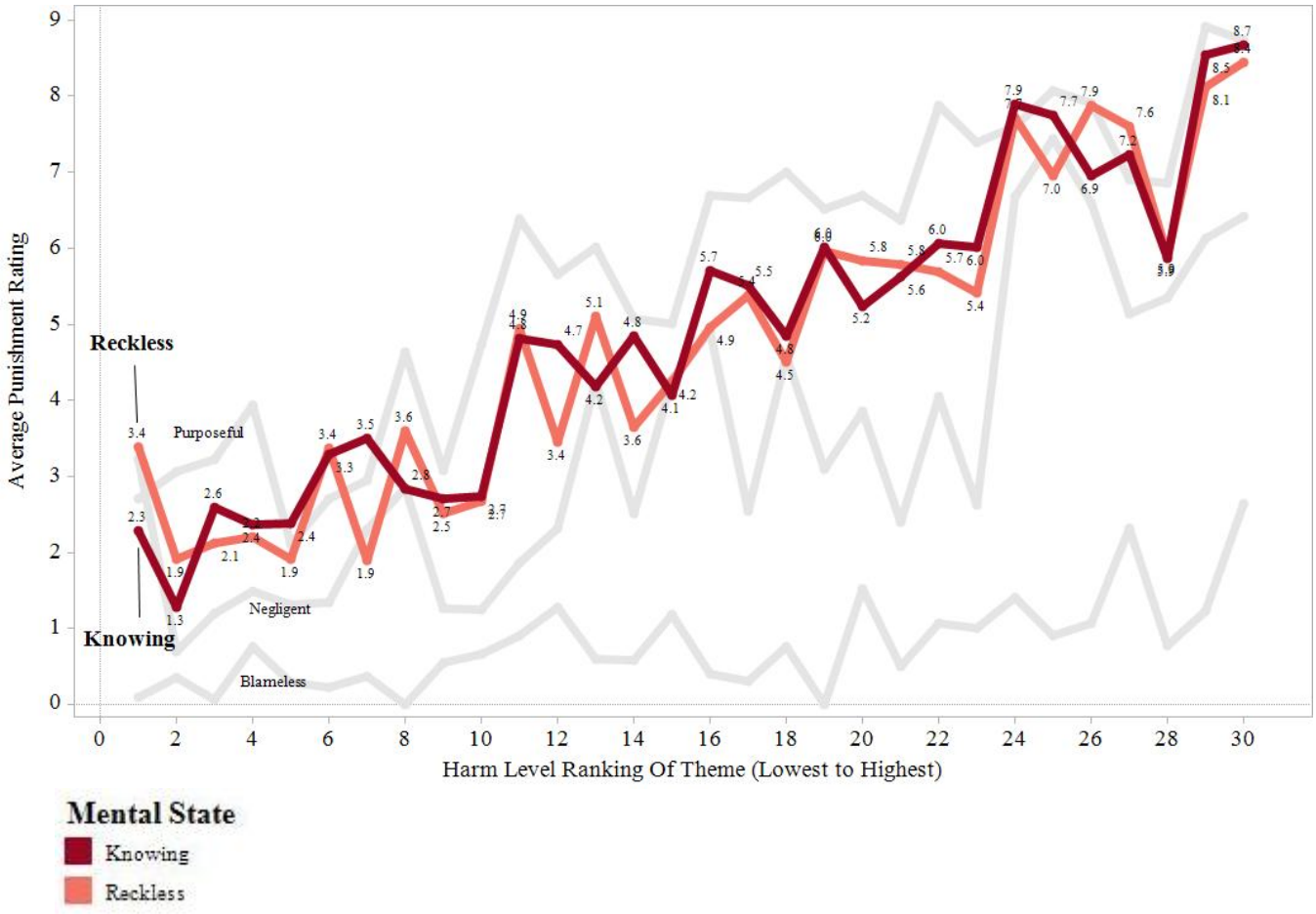
Figure 1. Average Punishment Ratings For Purposeful, Negligent, and Blameless Scenarios (Plotted By Harm Level Ranking of Theme)



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

Notes: Data for this figure is from Experiment 1, "Punishment Without Instructions". The y-axis plots average harm rating for each purposeful, negligent, and blameless scenario in each of thirty themes (averaged across all subjects who rated the particular scenario). Color shading indicates the mental state of the scenario.

Figure 2. Average Punishment Ratings For Knowing and Reckless Scenarios
(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, and far more than for any adjacent pairing. That is, they are *not* consistent with the assumptions of the Model Penal Code.

Notes: Data for this figure is from Experiment 1, "Punishment Without Instructions". The y-axis plots average harm rating for each knowing and reckless scenario in each of thirty themes (averaged across all subjects who rated the particular scenario). The dark red line indicates a knowing mental state, and the light red line indicates a reckless mental state. The three gray shaded lines, which are identical to the lines presented in Figure 1, are (from top to bottom) for purposeful, negligent, and blameless action, respectively.

The results in Figures 1 and 2, of course, were generated by subjects who did not have the benefit of 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 definitions once at the outset; in Experiment 3 we made the definitions available throughout – the confusion over K and R remains. To plot this flip-of-the-coin justice, we looked at K versus R in each of our thirty themes and in each of our four experiments (see Appendix A Figure A.1). We found 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. These results, presented and explained in the Appendix A, were verified through more detailed statistical analysis. There is no escaping the conclusion that our subjects cannot be readily trained to properly differentiate their punishment between K and R scenarios.

Why don't we see, even when we give subjects the MPC definitions, more differential between K and R? Two distinct possibilities present themselves. First, it could be the case 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 reckless act than he would assign to the same act done knowingly. If this is the case, then the problem is one of rating misalignment between the MPC and subjects' inherent culpability scorecard. A second possibility, however, is that the subject can't, from the start, tell the difference between recklessness and knowingly. In this second case, the subject would punish differently *if* he could sort properly. The problem is not one of rating, but of sorting ability.[62]

In order to distinguish between these two possibilities, we ran two additional experiments. First, to gather baseline data on sorting ability, in Experiment 4 subjects read through 30 scenarios. In addition to the scenario text, subjects were also provided with the definitions of the mental states. After reading each scenario,

[62] And, of course, it could be a combination of both of these explanations.

subjects were instructed: "Please select from the question options below the definition that best matches John's mental state in this scenario." We looked at how well subjects sorted, both whether they got it correct and, if they missed, how far away they were.[63]

We then extended Experiment 4 by combining our sorting and rating tasks into a new Experiment 5. In Experiment 5 subjects were first instructed to sort fifteen questions according to the MPC definitions (i.e. identical to the type of question in Experiment 4). After the sorting questions, however, subjects were given fifteen rating questions (these rating questions were presented without MPC definitions, as they were in Experiment 1).[64] Experiment 5 thus allowed us to line up individual punishment ratings with individual subject sorting ability. We were able to test whether "good sorters" punished differently than did "bad sorters".

The results of Experiment 4 suggest that subjects can identify, with a high degree of accuracy, purposeful and blameless scenarios. Looking at the average level of inaccuracy (as measured by the average absolute distance from the correct answer), we found that subjects were most prone to error in the middle categories (see Appendix A, Tables A.2 and A.1).

Looking more closely at the data, subjects can recognize well above chance, albeit with less accuracy than P and B, negligent action. Subjects do not, however, perform nearly as well when asked to identify knowing or reckless scenarios. In short, when dealing with knowing or with reckless scenarios, subjects are able to identity the correct mental state less than 50% of the time.[65]

---

[63] To do this, we coded two variables to measure subjects' sorting ability: (1) a dichotomous variable denoting whether subjects chose the correct category (correct = 1, incorrect = 0); and (2) a categorical variable, ranging from 0-4, measuring the absolute distance the subject was from the correct answer (i.e. 0 = correct answer, 1 = one category away, etc.). These two outcome variables allowed us to measure average sorting accuracy.

[64] 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.

[65] Note that, because there are five possible mental states to choose from, random guessing would produce a successful sorting rate of 20%. Thus,

These results suggest that subjects don't differentiate in punishment between K and R because they can't tell K and R apart. [66]

The final experiment we ran explored whether better mental states sorting would lead to sharper differentiation between the knowing and reckless punishment ratings. In the aggregate, we do not find that this is the case. Once again, the ratings for knowing and reckless cluster together, sitting lower than purposeful (see Figure 3 for the K vs. R comparison). Negligent is below the K/R cluster, and blameless is at the bottom. But what about the "good" sorters? When we re-run the analysis 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 knowing / reckless distinction is again blurred. Thus, while it may be that those who understand and can utilize the definitions (i.e. can sort correctly) punish slightly differently for K than for R, even the 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? First, caution is in order. Although we used a large number of subjects, this is still just one set of experiments in a young—indeed almost non-existent—empirical literature. Future studies will no doubt take us in many different, and some unanticipated, directions.

There are also some ecological and validation issues. For example, if it is true that people generally have difficulty seeing the differences between knowing and reckless mental states, then we, as the drafters of these scenarios, may also have had that

---

despite sorting accuracy of less than 50%, it is still above chance[so why using 50% criterion here?].

[66] When accuracy is measured by the absolute distance metric, the same pattern presents itself: subjects do best with the blameless and purposeful scenarios, and perform worst on knowing and reckless.

difficulty. Perhaps there are important situations in which our subjects could perceive a difference between knowing and reckless, but we just did a poor job of finding those situations. The successful results of our external validation with criminal law professors diminishes the likelihood that is true (though criminal law professors are admittedly trained to perceive these differences).

A related problem is that although our numbers 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 state by using different signaling language specifically designed to be conclusory about 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." 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. Yet if our signaling language artificially over-led subjects to the desired category, that would make the findings at the K-R distinction even *more* troublesome (and would make our subjects' abilities to distinguish all the *other* categories correspondingly less impressive).

Finally, there is the beguiling problem that the descriptive is not necessarily the prescriptive.[67] Even if it turns out to be true that humans can distinguish P, K/R, N and blameless states of mind from one another, and in assigning blame to those different states as the MPC suggests, that does not necessarily answer the

[67] This is what the moral philosophers call "the naturalistic fallacy," or sometimes also "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 LAW AND NEUROSCIENCE 483 (M. Freeman, ed., Oxford 2011). For contrasting views on whether this kind of moral realism might inform our punishment theories and practices, compare Donald Braman, Dan Kahan & David Hoffman, *Some Realism about Punishment Naturalism*, 77 U. CHI. L. REV. 1531 (2010) (criticizing punishment naturalism) with Paul Robinson, Rob Kurzban & Owen Jones, *Realism, Punishment and Reform*, forthcoming in U. CHI. L. REV. (defending punishment naturalism).

policy question of whether these "natural" categories should continue to be given legal traction. Indeed, there are many examples where legal systems have chosen to depart from these natural categories for policy reasons. The increasing list of strict liability regulatory crimes is one example. Drunk driving is another.

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. They may still serve important policies, some of which are discussed below. Perhaps we just need to find different ways of articulating the K/R difference, both in our legislation and in our jury instructions (though Experiments No. 2 and 3 suggest that if language is the problem, it is the legislative language and not the way we traditionally turn that language into jury instructions). Some suggestions in this regard are also discussed below.

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 reform.

So what are the plausible policy lessons to be drawn from our results? First, the good news for the MPC is that it seems the empirical cup is more than half-full. Except for K/R, subjects were fairly good at recognizing, then appropriately distinguishing, all the other levels of culpability. Keep in mind too that subjects were able to accomplish this differentiation *even in the absence of MPC instructions* (Experiment 1). This finding surprised several of us, including those of us who wrote the scenarios and had great difficulty coming up with ones that distinguished the P/K boundary. We expected subjects to have much more difficulty at that boundary than at the K/R boundary. The fact that they had very few difficulties at the P/K boundary not only tends to validate the MPC in this regard, it tends to validate the moral philosophers who have been teaching us for a long time that there is a moral difference between desired harm and harmful side-effects.

These results should also go a long way toward answering the broadest critics of the MPC approach, who continue to claim that the MPC categories are over-determined if not wholly

fictitious.[68] It seems, quite the contrary, that ordinary people are exquisitely tuned to make complicated moral judgments based on categorically different kinds of culpability, and that three of the four categories the MPC relies on are in the fact the categories that people use in the absence of a formalized code. The drafters of the MPC were right: we really do blame acts of desired harm (P) more than harmful side effects of desire (K); and we really do blame conscious and unreasonable risk-taking (R) more than inadvertent and unreasonable risk-taking (N). None of this should be surprising to students of evolutionary psychology. It makes evolutionary sense for an intensely social animal like us to use blame as a heuristic, or psychological substitute, for difficult but evolutionarily critical judgments about what to do with group members who violate group rules. And these particular categories—P, K/R, and N—themselves arguably made evolutionary sense.[69]

But we cannot ignore our other important finding. Subjects were simply unable to distinguish K from R, even when told to

---

[68] Note 21 *supra*. Because confession is good for the soul, the judge co-author of this paper must confess that he has flirted with the "mind-reader" criticism, and indeed thought that this experiment would go a long way toward proving that intentionality is not only a seamless continuum, but that it is seamlessly bound to harm. He has suggested in the past that criminal trial results would be exactly the same if we replaced the four MPC culpability categories with the following kind of jury instruction: "On a scale of one to four, how bad was this defendant?" Morris B. Hoffman, Booker, *Pragmatism and the Moral Jury*, 13 GEO. MASON L. REV. 455, 474 & n. 75 (2005) He admits that such views cannot withstand these empirical results. Neither, probably, can the position of those who have argued that criminal liability should be based largely on intent rather than harm. See, e.g., Larry Alexander, *Crime and Culpability*, 5 J. CONTEMP. L. ISSUES 1 (1994); Sanford H. Kadish, *The Criminal Law and the Luck of the Draw*, 84 J. CRIM. L. & CRIMINOLOGY 679 (1994); Steven J. Schulhofer, *Harm and Punishment: A Critique of the Emphasis on the Results of Conduct in the Criminal Law*, 122 U. PA. L. REV. 1497 (1974).

[69] Desire-based violators pose a much bigger future risk to the group, and therefore need more punishment, for no other reason than that they will certainly have future desires and have already shown a willingness to harm another members to satisfy their desires. Likewise, violators who *consciously* put other members at unreasonable risk probably poses a bigger future danger to the group than violators who took a risk without even knowing it, for no other reason than the latter kind of violator is easier to reform.

distinguish them and how to do so (by being given the definitions of K and R). This finding, if it ends up being validated in future studies, will demand some reforming attention.

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—murder—where the distinction is critical.[70] In almost every state, the sentencing differences between a knowing murder and a reckless one are enormous.[71]

In Colorado, for example, a Model Penal Code 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 16 years and 48 years.[72] A reckless murder, by contrast, is called manslaughter, and carries a non-mandatory sentence of 2 to 6 years.[73] In the very worst case,

---

[70] Only a handful of non-homicide MPC crimes require knowing conduct, and are not proved with even the most reckless conduct: arson (MPC §220.1), false imprisonment (§ 212.3) and sexual assault (§213.4). Perhaps the paucity of K/R crimes—and by that we mean 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.

[71] Only eight states—Delaware, Minnesota, Mississippi, New York, Oklahoma, Oregon, Washington and Wisconsin—follow the English model and do not use the K/R boundary to distinguish between levels of murder. Every other state either uses the K/R distinction as an express distinction in the definition of levels of murder, or effectively does so by making K an aggravator to an R murder.

[72] If the murder were purposeful (Colorado says "intentional") and after deliberation, it would be first degree murder, which is classified as a class 1 felony. COLO. REV. STAT. ANN § 18-3-102. CLASS 1 FELONIES CARRY A SENTENCE OF DEATH OR MANDATORY LIFE WITHOUT THE POSSIBILITY OF PAROLE. COLO. REV. STAT. ANN. §18-1.3-401(1)(a)(I) & (II). Second degree murder, without any heat of passion mitigator, is defined and classified as a class 2 felony at COLO. REV. STAT. ANN. §§ 18-3-103(1) & -103(3)(a). Class 2 felonies ordinarily carry a non-mandatory presumptive sentence of 8 to 24 years. COLO. REV. STAT. ANN. § 18-1.3-406(2)(a)(V)(A). But murder is also a crime of violence, which has the effects of increasing the range to 16 to 48 years, and making a prison sentence mandatory. COLO. REV. STAT. ANN. § 18-1.3-406(1)(a).

[73] Manslaughter is defined and classified as a class 4 felony in COLO. REV. STAT. ANN. § 18-3-104. Class 4 felonies carry a non-mandatory presumptive sentence of between 2 and 6 years. COLO. REV. STAT. ANN § 18-

therefore, the difference between a jury finding a knowing murder and a reckless one is the difference between a 48-year prison sentence and probation. The very *smallest* this difference could ever be is 10 years—the difference between the minimum of 16 years for a knowing murder and the maximum of 6 years for a reckless murder.

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."[74] In most states, the prosecution has the same right. This means that in virtually every case where a knowing murder is charged, and even sometimes when a purposeful murder is charged, the jury will also be asked whether the murder was merely reckless.[75]

It would therefore be a scandal of gigantic proportions if these differing sentences are the product of jurors making distinctions they simply cannot make. In our experiments, we included two themes in which a victim was killed by John.[76] Looking at the K and R sorting in these two themes, subjects were only correct about 50% of the time. What we might do about such

---

1.3-401(1)(a)(V)(A). Manslaughter is not one of the defined a crimes of violence under § 18-1.3-406.

[74] States differ on how they define a "lesser included" offense. There are two principal tests. Some states use the so-called "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). But more use the so-called "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? See generally, Christen R. Blair, *Constitutional Limitations on the Lesser Included Offense Doctrine*, 21 AM. CRIM. L. REV. 445 (1984).

[75] Of course, the vast 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.

[76] 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.

a scandal depends on the nature of the confound, which could have several explanations.

It may be that people do not view K primarily as a desire-based wrong but rather as a risk-taking wrong. Under this view, what is wrong about me 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. This explanation not only accounts for why subjects could not distinguish K from R (because both are about risk-taking, and subjects 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 that there is substantial risk of harm, then we desired the harm. This alternative would explain why subjects could robustly distinguish R from N (because the latter is a risk-taking wrong, and the former, in this account, a desire-based wrong). But we are not really in dire need of an alternative explanation for the R/N junction; the difference seems palpable. Reckless actors are conscious of the risk they are taking; 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.[77]

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

[77] See note 21 *supra*.

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."

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 helps, perhaps we should consider abolishing the distinction between K and R, at least in those states where that distinction takes on its most problematical form in the degrees of murder. If jurors cannot really tell these two categories apart, then at worst we are subjecting similarly-situated murder defendants to the vagaries of a meaningless distinction, with serious consequences. At best, we are inviting jurors to compromise their verdicts. In any event, the law risks legitimacy by imposing such serious consequences on a determination that seems to have no moral traction with its citizens.

On the other hand, there are several arguments for retaining the K/R difference even in the face of our results here, and even if these results cannot be avoided by better definitions and/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 have a palatable alternative to either the harshness of a P verdict or the leniency of an N verdict.

Second, keeping R as a separate level of culpability may also insulate us from one of the deepest, and continuing, debates in the criminal law—the extent to which the state should criminalize merely negligent behavior. R gives us a way to cabin the state's apparently insatiable desire to criminalize bad judgment to circumstances of *really* bad judgment. Even in circumstances where a crime of negligence goes to the jury, perhaps having the higher category of reckless 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 necessarily be limited to only keeping or jettisoning K and R as separate *general* categories of culpability. Legislatures may well have perfectly good reasons to decide a particular kind of "knowing" crime is more blameworthy than a particular kind of "reckless" crime. But our results suggest that even on a crime-by-crime basis, especially with murder, if this distinction is to continue to matter to legislatures, either they, in their code definitions, or trial courts in their instructions, will have to do a better job of articulating it.

CONCLUSION

At 50 years old, it is time for rigorous empirical evaluation of the fundamental assumptions underlying the Model Penal Code. Arguably, the most fundamental of all its assumptions is that typical jurors either naturally do – or at least can when instructed – sort culpable mental states into the four MPC categories: purposeful; knowing; reckless; and negligent.

Our experiments suggest that this is only partly true. Most damningly, subjects correctly identified knowing action in the vignettes only 44% of the time, and their success rate in recognizing reckless action was only 36%, *even when repeatedly instructed on the distinction*. This confusion leads to punishment patterns that are a far cry from the MPC's expectations. If subjects employed the MPC as envisioned, knowing actors would be punished more than reckless actors in each of our thirty themes. Our data, however, show that these two states of mind frequently are punished in exactly the reverse sequence from what the MPC prescribes.

If the findings reported here are robust, and if we as a society value treating like defendants and mental states similarly (or at least non-arbitrarily), then we should seriously reconsider the wisdom of using a regime in which the middle two of four culpable mental states are essentially indistinguishable by jurors.

Failing to consider such reform leaves us with a criminal system in which – at the important choice between knowing and

reckless states of mind – the very crime of which one is convicted, and the sentence that one accordingly receives, might as well be determined by a coin toss.

APPENDIX A: TECHNICAL DETAILS

This Appendix provides additional detail on the research design employed in our study, as well as the statistical procedures used to analyze the data.

A Closer Look At Research Design

The underdeveloped empirical literature on juror decision making on mental states leaves fundamental questions unsatisfactorily answered.[78] This is the result, as we will discuss in detail subsequently, of deficiencies in research design and subject samples. But such limitations notwithstanding, the studies provide a starting point for our empirical inquiry.[79] Relevant studies to date, of which there appear to be exactly four, employ a "scenario study" experimental approach.[80] Scenario studies can vary in

[78] In a 2005 study on mental states and the criminal law, law professor Justin Levinson observed similarly that an impressive array of scholarship points to the nuances and importance of mental state evaluations, but that "no psychologists have tested how jurors actually make *mens rea* judgments." Justin D. Levinson, *Mentally Misguided: How State Of Mind Inquiries Ignore Psychological Reality And Overlook Cultural Differences*, 49 How. L.J. 1, 7 (2005).

[79] A related body of research by psychologist Norman Finkel (in part collaborating with Jennifer Groscup) focuses on jurors' "commonsense justice" intuitions. In a series of scenario studies looking at mistakes of law and mistakes of fact, Finkel and Groscup (1997) found that assignment of culpability varied with assessment of the actor's intent. In the authors' words, intent "remains the starting point, and often the final point, in commonsense justice's culpability analysis. Though the course does not run smooth or straight from beginning to end but contextually wends its way through related variables, those variables generally feed back into intent, through interaction effects, modifying, mitigating, and sometimes enhancing culpability." Norman J. Finkel & Jennifer L. Groscup, *When Mistakes Happen: Commonsense Rules Of Culpability*, 3 PSYCHOL. PUB. POL'Y & L. 65, 117 (1997). More generally on commonsense justice, see NORMAN J. FINKEL, COMMONSENSE JUSTICE: JURORS' NOTIONS OF THE LAW (1995).

[80] Psychologist John Darley uses this descriptive phrase, and also notes that a second paradigm sometimes employed is the "experimental game" approach. John M. Darley, *Citizens' Assignments of Punishments for Moral Transgressions: A Case Study in the Psychology of Punishment*, 8 OHIO ST. J. CRIM. L. 101 (2010). While the experimental game approach has been used to

design, but at core they all rely on experimental subjects reading short, written scenarios about a bad act being committed by a single actor. To facilitate discussion, we referred throughout the Article to the actor in these scenarios as being named "John". Once subjects read about John's actions in a given scenario, subjects are then asked to make a judgment about John's mental state, John's blameworthiness, or what punishment, if any, John should receive. Researchers can manipulate a number of variables in these experiments, including:

- *The number and type of scenarios.* In order to know if individuals react differently to the same act done with different mental states, researchers must start by determining how many, and what types, of scenario comparisons they will make. For instance, researchers must decide whether they will conduct a "between subjects" experiment (e.g. where researchers compare subject A's response to scenario 1 with subject B's response to the same scenario 1), or a "within subjects" experiment (e.g. where subjects compare subject A's response to scenario 1 with the same subject A's response to a different scenario 2).
- *The subject matter of the scenarios*. Researchers must determine, for each type of scenario that will be compared, what the actor does and how the actor's mental state is to be described in the text. For example, suppose a researcher wants to compare responses to a scenario where John acts recklessly vs. a scenario where John acts in the same way, but with a negligent state of mind. To put this research design into practice, the researcher has to develop a fact pattern that is idenetical in all respects except for the mental state element.

---

assess punishment decisions in a number of contexts, to our knowledge, it has not been used for the questions we investigate in this Article. See John M. Darley, *Morality in the Law: The Psychological Foundations of Citizens' Desires to Punish Transgressions,* 5 ANN. REV. L. & SOC. SCI. 1, 9–14 (2009).

Moreover, the researcher must avoid a fact pattern that introduces confounding factors into the equation.[81]

- *Exposure, before or during presentation of the scenarios, to instructions about mental states*. Researchers can manipulate the quantity and detail of information subjects receive about mental states, for instance prompting thoughts about *mens rea* generally, providing statutory definitions, or offering no explicit guidance.
- *The nature of questions posed to subjects.* After reading the scenarios, subjects are asked a question, or a series of questions. Researchers must think carefully about what scales to use to measure the responses of their subjects. Researchers have investigated one or both of two outcomes of interest: (1) sorting ability, i.e. how well subjects can evaluate statutorily defined criminal law mental states in a scenario; and (2) assignment of punishment, i.e. how subjects assign different punishment levels to identical actions carried out with different *mens rea*.[82]

Decisions in each of these important areas during the research design phase leads to the construction of an experimental protocol.

---

[81] This is a crucially important step, as previous research suggests that details such as the perceived moral worth of the actor and the action can affect juror judgments. Philosopher Thomas Nadelhoffer has shown how moral judgments about the actor involved may influence mental state assessment. His study, however, like others in this area did not provide individuals with MPC definitions. Thomas Nadelhoffer, *Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality*, 9 PHILOSOPHICAL EXPLORATIONS 203 (2006).

[82] As we will discuss in Section IV, the relationship between "blame" and "punishment" is not 1:1. See supra, note 60.

Running the Experiments

We ran the experiments in May 2010.[83] We used a web-based experimental platform, which allowed us to recruit a large and diverse sample of subjects.[84] We used the web survey firm Qualtrics to both recruit our subjects and host our web-based experiment.[85] Research using Qualtrics-based experiments has been published and presented in a number of academic fields, suggesting that it meets scholarly expectations for quality online web-based experiments.[86] All subjects recruited by Qualtrics were United States citizens, age 18 to 65. We further made this a jury-eligible subject pool by filtering out (via an initial screening question) subjects who indicated that they had been convicted of a felony. Qualtrics uses opt-in survey panels to recruit subjects. Subjects were recruited from the general population, and no personally identifying information was collected.

---

[83] A first round of experiments, run in March 2010, are not reported in the Article but produced results substantively similar to those we report. All experiments received approval from the University of California, Santa Barbara Institutional Review Board. NRII-SH-FR-019-1N.

[84] 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 Psychology 243 at 244. (2001).

[85] Subject costs were approximately $7 per subject for a nationally representative sample. Qualtrics recruits subjects only through its proprietary opt-in email lists. For more on Qualtrics, see: www.qualtrics.com.

[86] Studies relying on Qualtrics experiments include: Dan Simon Doug Stenstrom Stephen Read. Partisanship and Prosecutorial Decision Making: An Experiment. Working paper presented at the Nov. 2009 Conference on Empirical Legal Studies. Dannagal G. Young and Lindsay Hoffman. An Experimental Exploration of Political Knowledge Acquisition from the Daily Show Versus CNN Student News. Paper presented at the 2009 Meeting of the American Political Science Association. Jonathan S. Abramowitz, Gerald R. Lackey, Michael G. Wheaton.Obsessive–compulsive symptoms: The contribution of obsessional beliefs and experiential avoidance. Journal of Anxiety Disorders 23 (2009) 160–166. Yany Grégoire, Thomas M. Tripp, & Renaud Legoux. When Customer Love Turns into Lasting Hate: The Effects of Relationship Strength and Time on Customer Revenge and Avoidance. Journal of Marketing. Volume 73, Number 6, November 2009.

At the end of the experiment, we collected demographic information on subjects. This demographic data allows us to look at the representativeness of our sample (reported in Appendix Table A.1). While not a truly nationally representative sample, the 529 subjects who participated in the experiments we report in this Article came from 48 states. Our sample was roughly equal in terms of gender, with 53% subjects female and 47% male. Our subjects were older, on average, than the comparable U.S. population. Our sample was 88% white, higher than the national average. In terms of education, our subjects are slightly skewed toward having more education. Income distributions of our subjects and the U.S. population as a whole are similar, though not entirely the same. Clearly our sample, both in its size and demographic makeup, is more reflective of a jury pool than any of the previous studies discussed in Part II above.

Table A.1. Demographics of Experimental Subjects (N = 529)

Education	**Subjects**	**U.S. Census**
Less than HS	1%	18%
High school / GED	21%	30%
Some college	33%	20%
Assoc. degree	12%	7%
Bachelor's	22%	17%
Graduate Degree	11%	10%

Income	**Subjects**	**U.S. Census**
< $20,000	18%	$1 to $24,999: 22%
$20,000 - $40,000	33%	$25,000 to $34,999: 19%
$40,000 - $60,000	20%	$35,000 to $49,999: 21%
$60,000 - $80,000	13%	$50,000 to $64,999: 14%
$80,000 - $100,000	7%	$65,000 to $74,999: 6%
> $100,000	8%	$75,000 to $99,999: 8%

Gender	**Subjects**	**U.S. Census**
Male	47%	49%
Female	53%	51%

Jury member in criminal case?	**Subjects**
Yes	18%
No	82%

Age Groups	**Subjects**	**U.S. Census**
18-24	5%	13%
25-34	12%	18%
35-44	19%	19%
45-59	46%	27%
60 +	18%	23%

Race	**Subjects**	**U.S. Census**
White	88%	74%
Non-White	12%	26%

Concerns about internal validity – whether subjects are actually doing what you've instructed them to do – are always a concern in experiments, but are especially important to address with online experiments because subjects cannot be monitored while engaged in the experimental tasks.[87] To address this issue, experimental psychologists have developed "attention filters" designed to ascertain whether subjects are in fact paying attention to the material being presented to them online. These experiments employed a modified version of the filter developed by psychologist Danny Oppenheimer and colleagues.[88] The design of the attention filter question was such that users who did not read carefully would see, in large font the 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 (e.g. Local newspaper, local TV news, etc). Subjects reading carefully, however, were instructed *not* to check any of the boxes, but instead to type the numbers 123 into the text box provided.[89] All of the results presented in this Article are based only on those "good" subjects, i.e. those subjects who were paying attention.

Details of Confirmatory Statistical Analysis

In the body of the Article we present a series of graphical figures. Here we present the statistical analysis confirming that the graphical patterns we see with the naked eye are in fact statistically significant patterns. Looking first at punishment ratings in

[87] A filter employed after data collection allowed for the experiment to exclude from the dataset subjects with duplicate IP addresses.

[88] Daniel M. Oppenheimer, Tom Meyvis, & Nicolas Davidenk, *Instructional Manipulation Checks: Detecting Satisficing To Increase Statistical Power*, 45 Journal of Experimental Social Psychology 867 (2009).

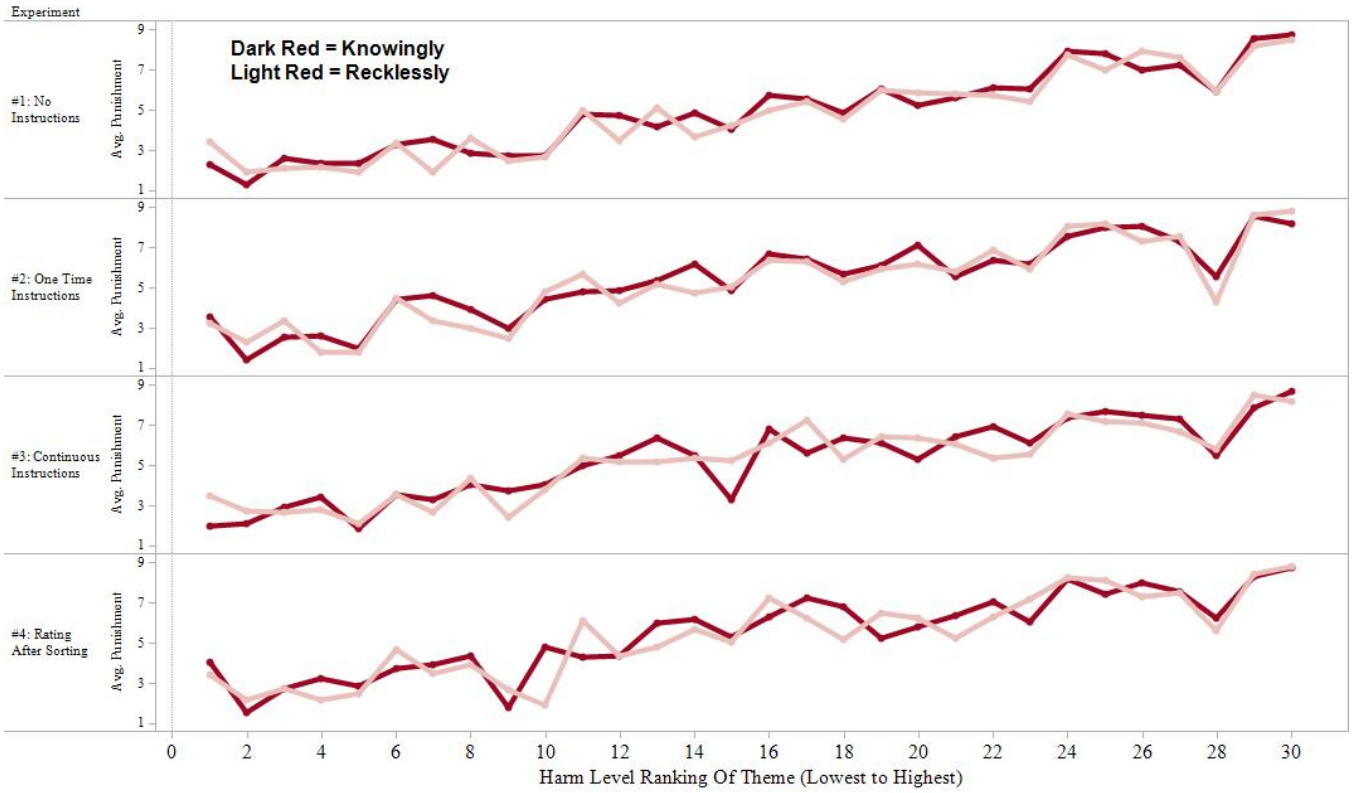
[89] Across the five experiments, 45% of subjects successfully answered the attention filter question. Additional analysis suggests that even when including the "bad" responses, substantive effects do not differ sharply. Moreover, a question for speculation is whether we imagine citizens and jurors to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

Experiment 1, we ran a straightforward ordinary least squares (OLS) regression of standardized punishment on mental state. The regression model employs clustered (on subject) robust standard errors, to account for the fact that punishment ratings are not independent but grouped by subject. Post-estimation tests confirm our intuitions from Figures 1 and 2. There is a statistically significant difference in mean standardized punishment rating between purposeful and knowing ($F(1, 90) = 58.21$, $p<.01$), but not between knowing and reckless ($F(1, 90) = 2.52$). Reckless is significantly different from negligent ($F(1, 90) = 118.50$, $p<.01$), and there is a significant difference between negligent and baseline blameless ($F(1, 90) = 688.04$, $p<.01$). Substantively similar results were also produced by an alternative model specification employing punishment ratings as the dependent variable, and introducing theme-specific random effects introduced.[90]

In Experiment 2, as in Experiment 1, there is a statistically significant difference in mean standardized punishment rating between purposeful and knowing ($F(1, 94) = 119.07$, $p<.01$), but not between knowing and reckless ($F(1, 94) = 2.71$). Reckless is significantly different from negligent ($F(1, 94) = 112.60$, $p<.01$), and there is a significant difference between negligent and baseline blameless ($F(1, 94) = 576.97$, $p<.01$). In Experiment 3, there is a statistically significant difference in mean standardized punishment rating between purposeful and knowing. But there remains no significant difference between knowing and reckless. Summarizing graphically across experiments 1 through 4 in Figure A.1, we can see that – regardless of our instructions – the K and R ratings were not distinct and frequently were reversed from MPC expectations.

[90] 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 subject's 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. Of the subject demographic variables, only age was statistically significantly related to punishment, with older subjects assigning more punishment.

Figure A.1. Comparing Average Punishment Ratings For Knowing And For Reckless Scenarios, Across Four Different Experimental Designs



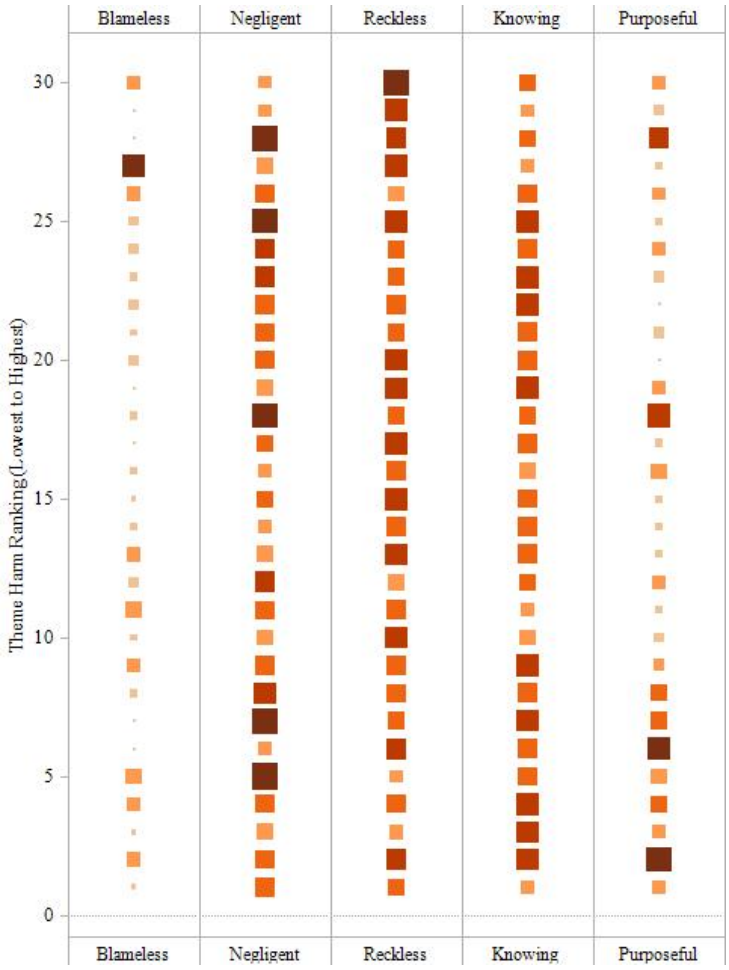
What To Notice In Figure A.1: Regardless of the experimental design, the average punishment ratings for knowing and reckless scenarios were very similar, and they frequently reversed. The figure illustrates the point that – whether subjects had no instructions, one-time instructions, continuous instructions, or rated after sorting – subjects did not clearly and regularly rate knowing behavior as more punishment-worthy than reckless behavior.

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

Turning to the sorting experiments, the results of which are presented in Appendix Figure A.2 and Table A.2, we ran a logit regression of the correct/incorrect variable on four dichotomous variables for purposeful, knowing, reckless, and negligent. Blameless 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-2 tests confirm that there is a statistically significant difference in the likelihood of correctly sorting between purposeful and knowing ($chi2(1) = 61.32$, $p<.01$), between knowing and reckless ($chi2(1) = 10.17$, $p<.05$), between reckless and negligent ($chi2(1) = 7.92$, $p<.05$), and between negligent and blameless ($chi2(1) = 139.79$, $p<.01$). There is no significant difference between negligent and knowing ($chi2(1) = 0.39$).

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. All regression models employ clustered (on subject) robust standard errors, as discussed previously. Added to the logit model described above were a standardized measure of theme harm level, and measures of subject's age, race, political ideology, education level, and past experience as a crime victim. In this expanded model, all relationships reported above still hold.

Figure A.2. Sorting Difficulty In Experiment 4, By Theme and By Mental State



What To Notice In Figure A.2: As indicated by the larger, darker shaded boxes, subjects had the greatest difficulty in correctly identifying scenarios in the middle categories of Negligent, Reckless, and Knowing.

Notes: Figure A.2 is a graph of the average absolute distance from correct answer, by mental state and by theme. Darker colors and larger sizes of boxes on the graph indicate greater average distance from the correct answer. The smaller and lighter boxes indicate more sorting accuracy. Data for this graph is from Experiment 4 ("Sorting").

Table A.2 Sorting Success Rate In Experiment 4, By Mental State

	Correct Mental State:				
	Purposeful	Knowing	Reckless	Negligent	Blameless
What Subjects Chose:					
Purposeful	**79%**	12%	5%	3%	1%
Knowing	11%	**44%**	27%	14%	1%
Reckless	7%	37%	**36%**	15%	1%
Negligent	3%	5%	31%	**57%**	13%
Blameless	0%	1%	2%	11%	**84%**

What to Notice in Table A.2: The success of subjects in recognizing a knowing scenario as a knowing scenario, or in recognizing a reckless scenario as a reckless scenario, is below 50% (44% and 36% respectively), whereas for each of the others the success rate is above, or much above, 50%.

Note: The gray cells in Table 3 display the sorting success rate for each mental state. The non-gray cells display the percentage of subjects across the other four (and incorrect) options. For instance, looking at the column labeled "Purposeful", 79% of subjects correctly identified these scenarios; 11% mistook them for knowing; 7% mistook them for reckless; 3% mistook them for negligent; and 0% mistook them for blameless.

To examine accuracy as measured by absolute distance, we employed an ordered logit model. Post-estimation chi-2 tests confirm that there is a statistically significant difference in the absolute magnitude of sorting error between purposeful and knowing ($chi2(1) = 7.54$, $p<.01$), and between knowing and reckless ($chi2(1) = 220.6$), though not between knowing and negligent ($chi2(1) = 2.52$). There is a significant difference between negligent and blameless ($chi2(1) = 23.74$, $p<.01$). Taken together, these findings on absolute difference are consistent with the analysis of correct sorting: the extremes (purposeful and blameless) generate less error than do the middle three categories (knowing, reckless, negligent). Expanded regression models,

adding in controls for harm level and subject demographics, did not affect these substantive results.

Finally, and mirroring the results in Experiment 1, in Experiment 5 there is a statistically significant difference in mean standardized punishment rating between purposeful and knowing ($F(1, 148) = 70.65$, $p<.01$), but not between knowing and reckless ($F(1, 148) = 0.85$). Reckless is significantly different from negligent ($F(1, 148) = 109.59$, $p<.01$), and there is a significant difference between negligent and baseline blameless ($F(1, 148) = 878.25$, $p<.01$).

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 on-line at this location: < http://law.vanderbilt.edu/download.aspx?id=6421 >.