Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability

Erik Lillquist[*]

According to the traditional understanding, proof beyond a reasonable doubt requires a high level of certainty before a jury convicts someone of a crime. This view is often framed in explicitly utilitarian terms: that a high standard of proof is justified because the costs of erroneously convicting an innocent person are so much higher than the costs of erroneously convicting a guilty person. Empirical evidence, though, suggests that the standard reasonable doubt jury instruction does not actually require as much certainty as we generally assume. The common response to this observation is to propose improvements to the instruction to ensure that juries will require a high level of certainty. This Article contends that this solution is misguided because it wrongly presumes that the instruction needs to be fixed. Instead, what needs to change is our understanding of reasonable doubt. Using the expected utility model of decision theory, as well as insights from behavioral economics and the social norms literature, this Article suggests that the reasonable doubt standard of proof is inevitably flexible in nature: in some cases juries will require more proof than in other cases. The Article goes on to suggest that this result is in fact preferable to a standard of proof that requires a high level of certainty in all criminal cases.

---

[*] Associate Professor of Law, Seton Hall University School of Law. B.S. & B.A., Stanford University; J.D., University of Virginia. I would like to thank Michelle Adams, Ronald Allen, Jake Barnes, Kathleen Boozang, Neil Cohen, John Jacobi, Jennifer Mnookin, Michael Risinger, Daniel Solove, William Stuntz, Charles Sullivan, Heather Taylor and Sarah Waldeck for their comments on previous drafts. I also received many useful comments from other faculty members at Seton Hall when I presented an embryonic version of this article in the Fall of 1998, from the participants at the 2002 Stanford/Yale Junior Faculty Forum (particularly Dan Kahan, Stephen Morse and Mitchell Berman), and from the participants at a workshop at the University of Virginia School of Law. I would also like to thank Marita Erbeck, Candice Chesson, and Kate McNamara for their valuable research assistance.

TABLE OF CONTENTS

> *[T]he law holds, that it is better that ten guilty persons escape, than that one innocent suffer.*[1]

> *The old ACLU notion that it's better to let 10 guilty men go free than to imprison one innocent person is, in an age of suicide bombers with access to bioweapons, not just a luxury but a danger.*[2]

What is the quantum of proof necessary to convict a person of a crime in America? The standard assumption — by the public, by lawyers, and by academics — is that we require a very high certainty of guilt before

---

[1] 4 WILLIAM BLACKSTONE, COMMENTARIES 352 (1769).

[2] Jonathan Alter, *Keeping Order in the Courts*, NEWSWEEK, Dec. 10, 2001, at 48.

convicting anyone of a crime. We also generally assume that this high level of certainty is enforced through the requirement of proof beyond a reasonable doubt in all criminal cases, a standard that has governed in American criminal cases since the late 18th Century.[3] But what the standard really requires and why we use it at all both remain unclear.

Consider what "proof beyond a reasonable doubt" actually mandates that the jury do. Surely it requires more proof than the preponderance of the evidence standard, which governs in civil cases. As commonly explained to civil juries, the preponderance standard is quantified as any amount of certainty greater than 50%, and proof beyond a reasonable doubt must mean more than that.[4] But how much more proof than a preponderance is needed in a criminal case? The quantity of certainty is never quantified; instead, it is kept quite vague. Is 90% certainty required? 95%? 99%? Or could the amount of certainty be much lower, say perhaps 75%?

Guidance on these questions certainly does not come from the reasonable doubt jury instructions. One of the most prominent federal pattern jury instruction treatises defines reasonable doubt as follows:

> The question naturally is what is a reasonable doubt. The words almost define themselves. It is a doubt based upon reason and common sense. It is a doubt that a reasonable person has after carefully weighing all of the evidence. It is a doubt which would cause a reasonable person to hesitate to act in a matter of importance in his or her personal life. Proof beyond a reasonable doubt must . . . be proof of such a convincing character that a reasonable person would not hesitate to rely and act upon it in the most important of his own affairs.[5]

The instruction talks about reasonable doubt as measured by the certainty a reasonable person would need to act without hesitation in a matter of importance. How high a level of certainty is that? Indeed, what level of certainty does such a person need? What counts as a matter of importance?

---

[3] *See* BARBARA J. SHAPIRO, BEYOND REASONABLE DOUBT AND PROBABLE CAUSE: HISTORICAL PERSPECTIVES ON THE ANGLO-AMERICAN LAW OF EVIDENCE 20-25 (1991); Anthony A. Morano, *A Reexamination of the Development of the Reasonable Doubt Rule*, 55 B.U. L. REV. 507, 519-24 (1975).

[4] 2 MCCORMICK ON EVIDENCE 428-29 (John W. Strong ed., 5th ed. 1999).

[5] 1 LEONARD B. SAND ET AL., MODERN FEDERAL JURY INSTRUCTIONS, Instr. 4-2, at p. 4-8 (2002).

In fact, reasonable doubt jury instructions tell the juror very little. They make no attempt to place a firm numerical value on the certainty that is required, and they can be read as simply suggesting that some indeterminate level of certainty is sufficient. That may be why the instructions require a level of certainty where a reasonable person would *hesitate* to act in a matter of importance. But nothing in the instructions requires the instructions to be understood that way and they are certainly capable of being interpreted by the jurors in other ways. Furthermore, as if to make matters worse, there is no way to enforce the instructions' command as to the proper standard of proof: the jury cannot be sanctioned in any way if it decides to ignore completely whatever guidance it has been given on the standard of proof.[6]

The intuition that the standard may not always require a high level of certainty is confirmed by empirical studies. Ever since the publication of Harry Kalven and Hans Zeisel's *The American Jury*[7] in 1966, researchers have been subjecting reasonable doubt, as well as other standards of proof, to empirical tests. This research has consistently shown that the jurors in criminal cases will often be satisfied with much less certainty than is conventionally assumed.[8]

Several commentators, in response to this evidence, have suggested various steps to increase its rigor. Justice Ginsburg and Professor Lawrence Solan, for instance, have both suggested that the solution is to adopt an instruction proposed by the Federal Judicial Center, which describes reasonable doubt as requiring that the jury be "firmly convinced" of the defendant's guilt.[9] Others have suggested different improvements to the instructions, such as more widespread use of pattern instructions, or even the possibility of giving *no* definition of reasonable doubt at all.[10]

---

[6] *See infra* notes 123-145 and accompanying text. This does not mean that jurors consciously disobey the instructions as a routine matter. *Cf.* E. Allan Lind, *The Psychology of Courtroom Procedure*, *in* THE PSYCHOLOGY OF THE COURTROOM 13, 30-31 (Norbert L. Kerr & Robert M. Bray eds. 1982) (noting that instructions do have some effect). Deviations do occur, though, either because of conscious decisions or, more likely, subconscious decisions. The point is just that if jurors consciously or subconsciously disregard the instruction, there is little the legal system does to correct this.

[7] HARRY KALVEN, JR. & HANS ZEISEL, THE AMERICAN JURY (1966).

[8] *See infra* Part II.

[9] Victor v. Nebraska, 511 U.S. 1, 26 (1994) (Ginsburg, J., concurring in part and concurring in judgment); Lawrence M. Solan, *Refocusing the Burden of Proof in Criminal Cases: Some Doubt About Reasonable Doubt*, 78 TEX. L. REV. 105, 144 (1999); *see also* Jon O. Newman, *Beyond "Reasonable Doubt,"* 68 N.Y.U. L. REV. 979, 991 (1993); Robert C. Power, *Reasonable and Other Doubts: The Problem of Jury Instructions*, 67 TENN. L. REV. 45, 121 (1999).

[10] Henry L. Chambers, Jr., *Reasonable Certainty and Reasonable Doubt*, 81 MARQ. L. REV.

Such proposals presume a fix is needed; they assume that the standard of proof in criminal cases *must* be high and that if, in practice, it is not high, then the system needs to be fixed. But why does the standard of proof in criminal cases have to be high? Why can it not be lower, at least in some cases? This Article contends a reasonable doubt standard that instead varies from case to case is, in fact, more beneficial to society.

The most common argument for a high level of certainty in all criminal cases, and the one that this Article will refer to as the traditional view, is that high certainty is required to protect innocent defendants against erroneous guilty verdicts.[11] The analysis is premised on an explicitly utilitarian trade-off: a high degree of certainty has the benefit of lowering the possibility that an innocent person will be convicted, at the cost that some guilty people will be free. The traditional view presumes that the reasonable doubt standard is high and then justifies it based on the calculus that it is more costly to erroneously convict an innocent person than it is to erroneously acquit a guilty person. This theory has been cited by those on both sides of the legal spectrum, from Justices Brennan and Harlan in *In re Winship*[12] to Judge Richard Posner and Professor Richard Epstein.[13]

655, 697-703 (1998) (discussing various possibilities); Jessica N. Cohen, *The Reasonable Doubt Jury Instruction: Giving Meaning to a Critical Concept*, 22 AM. J. CRIM. L. 677, 696-701 (1995) (arguing for Devitt & Blackmar instruction); Shelagh Kenney, Note, *Fifth Amendment — Upholding the Constitutional Merit of Misleading Reasonable Doubt Instructions*, 85 J. CRIM. L. & CRIMINOLOGY 989, 1025-26 (1995) (arguing for improvement in comprehensibility of reasonable doubt instructions); Note, *Reasonable Doubt: An Argument Against Definition*, 108 HARV. L. REV. 1955, 1967-68 (1995) [hereinafter *Reasonable Doubt*].

[11] *See* Daniel Shaviro, Commentary, *Statistical-Probability Evidence and the Appearance of Justice*, 103 HARV. L. REV. 530, 530 (1989) (citing Tribe, *infra* note 13, at 385 n.64). *See, e.g.*, *In re* Winship, 397 U.S. 358, 363-64 (1970); A.A.S. ZUCKERMAN, THE PRINCIPLES OF CRIMINAL EVIDENCE 123, 135 (1989); Richard J. Allen, *The Restoration of In re* Winship*: A Comment on Burdens of Persuasion in Criminal Cases after Patterson v. New York*, 76 MICH. L. REV. 30, 47 (1977) [hereinafter Allen, *Restoration*]; Chambers, *supra* note 10, at 656-57; Morano, *supra* note 3, at 507-09; Newman, *supra* note 9, at 981-85; Power, *supra* note 9, at 51-52; Solan, *supra* note 9, at 110-12; Henry A. Diamond, Note, *Reasonable Doubt: To Define, or Not to Define*, 90 COLUM. L. REV. 1716, 1717 (1990). Of course, the jurors themselves are not told of this weighing of errors costs.

[12] 397 U.S. at 363; *id.* at 372 (Harlan, J., concurring); *see also* Ballew v. Georgia, 435 U.S. 223, 232-38 (1978) (discussing study that weighed costs of erroneous acquittals and convictions in arriving at optimal size for criminal jury); Scott E. Sundby, *The Reasonable Doubt Rule and the Meaning of Innocence*, 40 HASTINGS L.J. 457, 458, 459-61, 467 (1989).

[13] RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 604-05 (5th ed. 1998) [hereinafter POSNER, ECONOMIC ANALYSIS]; RICHARD A. EPSTEIN, FORBIDDEN GROUNDS: THE CASE AGAINST EMPLOYMENT DISCRIMINATION LAWS 225 (1992); Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399, 408-17 (1973) [Posner, *Economic Approach to Legal Procedure*]; *see also* Frederick Schauer & Richard Zeckhauser, *On the Degree of Confidence for Adverse Decisions*, 25 J. LEGAL STUD. 27, 34 (1996)

As far as it goes, the traditional understanding of reasonable doubt is consistent with the expected utility model that dominates decision theory, as well as much of modern law and economics.[14] Under the expected utility model, when a rational person is making a decision under uncertainty, she should pick the option that has the greatest expected utility. The expected utility of a particular option is the sum of the utility of the possible outcomes after the decision, weighed by the probability of each possible outcome.[15] Thus, in a criminal case, the expected utility of a decision to convict a defendant is the utility of an accurate conviction — weighted by the probability the defendant is in fact guilty — plus the disutility of an erroneous conviction — weighted by the probability that the defendant is in fact not guilty.

If, as Richard Posner argues, "the net social cost of erroneous conviction is high [and t]he net social cost of acquitting a guilty person is, in contrast, apt to be low,"[16] then it makes sense to set the standard of proof at a level that will have fewer erroneous convictions than acquittals. Using expected utility theory, it is possible to show that if, for instance, erroneous convictions are ten times more costly than erroneous acquittals (as is often assumed), then the standard of proof in a criminal case should be about 90%.

This Article contends that this explanation is fundamentally flawed, because it fails to account properly for the costs and benefits of convictions and acquittals in criminal cases. First, the model set forth by the traditional view is too simplistic. It is not enough to weigh only the

---

(stating that choice of standard of proof "is an exercise in trading off the harms that flow from different types of error"); 2 MCCORMICK ON EVIDENCE, *supra* note 4, at 517 n.2 (citing *In re* Winship). English thinkers have also continued to use the traditional theory. *See* ZUCKERMAN, *supra* note 11, at 123.

Some have rejected the traditional view and suggested instead that the reasonable doubt standard is best explained not by focusing on the level of certainty that the standard requires, but instead by looking to the legitimacy that the standard creates for the legal system. *See, e.g.*, Laurence H. Tribe, *Trial By Mathematics: Precision & Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1375 (1971) [hereinafter Tribe, *Trial by Mathematics*] (arguing existence of compromise between not insisting on guilt where less than certain and high cost of being explicitly precise); Charles R. Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 HARV. L. REV. 1187 (1979) [hereinafter Nesson, *Reasonable Doubt and Permissive Inferences*]. This Article addresses these theories in Part IV.C, *infra*.

[14] *See, e.g.*, POSNER, ECONOMIC ANALYSIS, *supra* note 13, at 12-17; ROBERT COOTER & THOMAS ULEN, LAW & ECONOMICS 42-48 (2d ed. 1997).

[15] *See* Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, 47 ECONOMETRICA 263, 265 (1979) (noting that in expected utility theory "utilities of outcomes are weighted by their probabilities").

[16] POSNER, ECONOMIC ANALYSIS, *supra* note 13, at 605.

costs of erroneous convictions and acquittals; *accurate* convictions must be weighed in reaching any decision about the proper standard of proof. Factoring these benefits into the mix substantially alters where the standard of proof should be set.

Second, the traditional understanding wrongly assumes that we know *how* to weigh these costs and benefits. Using insights from behavioral economics and the social norms literature, this Article demonstrates that there is no existing theory that can explain what the relative costs and benefits of the various verdict possibilities ought to be across all cases. Instead, the costs of erroneous verdicts and the benefits of accurate verdicts are certain to vary depending upon both the nature of the crime and the character of the defendant.

Drawing on these critiques of the traditional expected utility models, I suggest that the standard of proof in criminal cases *should* vary from case to case. Not only is this consistent with the way juries appear to implement the standard in practice, but it is a socially preferable outcome. It is also a result that can work to benefit defendants, particularly in serious cases. For instance, in cases involving the death penalty, a flexible standard of proof should usually require nearly absolute certainty: a standard of proof that is above even what we commonly assume to constitute reasonable doubt. In practice, this standard may not be that high, but this is the result of Supreme Court cases that allow the government to exclude jurors who would otherwise drive up the standard of proof.[17] This theory, therefore, suggests that such juror exclusion rules have the undesirable effect of lowering the standard of proof in capital cases below the optimal level.

It also easy to imagine other criminal cases where the standard of proof will be lower. For instance, in a case involving someone accused of terrorism and known to be an outspoken advocate of such actions, the standard of proof will probably be lower than in other cases, because the risk of harm from an erroneous acquittal is higher than in other cases. This Article contends that such a result is preferable, as a proper balancing of the costs and benefits to society justifies the lower standard of proof. Similarly, there may also be cases involving low-level "criminality," such as traffic offenses or other nuisances, where the disutility of an erroneous conviction is comparatively low. In such cases, it also makes sense to have a lower standard of proof.

Part I begins by explaining the interaction between the standard of proof and errors and then sets forth the traditional theory and, in

---

[17] *See infra* note 214.

particular, its formalization over the past thirty-five years. This Part refines the model by including not only inaccurate verdicts, but also accurate verdicts. Part II then shows that there is little reason to suppose that jurors apply a high fixed standard of proof in all criminal cases. First, empirical evidence suggests that jurors do not apply anything close to the level of certainty that is typically thought to be required by reasonable doubt. Second, as a theoretical matter, jurors themselves will apply the standard that they themselves prefer, rather than the standard we assume they receive in their instructions, and there are strong reasons to believe that jurors themselves should prefer a flexible standard.

Part III explores another conception of the traditional understanding. It is a normative model of how jurors *ought* to set the standard of proof, rather than a description of how jurors do set the standard of proof. On this view, the standard of proof in a criminal case ought to be high because of the relative utilities of erroneous and accurate verdicts. Even when the traditional understanding is analyzed this way, though, it still fails. The difficulty is that the decision theory model itself cannot provide the underlying utilities and existing theories of criminal law fail to generate utilities that suggest that the standard of proof should be high in all cases.

Part IV sets forth an alternative theory of reasonable doubt. Like the traditional view, this flexible theory views the reasonable doubt standard as attempting to strike a balance between the utilities and disutilities associated with errors. This new theory, though, draws on the insight that no agreement can be reached concerning the proper normative value for the reasonable doubt standard. The law appropriately leaves the amount of certainty unstated, so that the jurors can make the final decision on the proper level. This is because the optimal level of certainty in any particular case is both crime- and defendant-specific — some particular types of crime and some particular defendants might optimally require very high levels of certainty, while other crimes and defendants might optimally require much lower levels of certainty. The use of a flexible standard allows the decision maker to apply the level of certainty that is most appropriate to a particular case.

Part IV then goes on to demonstrate that a flexible standard is preferable to a rule that applies a consistent standard to be used across all cases. Of course, both approaches — the flexible standard and the more rigid rule — would create possible errors. In general, however, fewer and less costly errors are made by having a flexible standard. This Part shows that the reasonable doubt standard relies upon the actual decision maker (either a jury or a judge) coming to its own rough

determination of the proper level of certainty in a particular case, and that these resolutions are more likely to be in accord with our view of a particular case than would a result arrived at by using a rigid rule of reasonable doubt. The Part then concludes by responding to those commentators who have suggested that the need for legitimacy in the criminal justice system, and not decision theory, explains the standard of proof.

Part V turns to why the standard is not described to the jury in a way that would encourage the jury to vary the standard from case to case. In other words, to the question of why, if we want jurors to vary the standard, do we not use a more precise instruction? Part V shows that the present vague reasonable doubt instruction can be explained as facilitating the use of a variable standard. Attempts to define reasonable doubt more specifically would be unstable and, far from encouraging the use of a variable standard, would inevitably lead to attempts to make the standard more rigid.

I. REASONABLE DOUBT: THE TRADITIONAL VIEW

Cases require decisions, both about the law and the facts. Sometimes those decisions concern solely the law and are made by judges. Frequently, however, decisions need to be made about the facts themselves. In the Anglo-American legal system, decisions about facts — particularly in criminal cases — are generally made by the jury.[18]

One way to ensure that juries never make mistakes would be to require complete certainty — we would never allow a jury to return a verdict, either guilty or not guilty (in a civil case, liable or not liable), until the jury was absolutely convinced. We routinely assume, however, that the jury can rarely be entirely confident in deciding what those facts are, because attempts to reconstruct past events are inherently uncertain. So instead, we require that one of the parties convince the decision maker of the particular fact to a designated level of certainty, called the standard of proof. In civil cases, this standard of proof is the

[18] There are, of course, many exceptions to this vast overgeneralization. For instance, the parties can agree to forego a jury decision on the facts in favor of a judge. *See* FED. R. CRIM. P. 23(a) (requiring that in federal criminal cases, jury trial can be waived "with the approval of the court and the consent of the government"). A jurisdiction can require in a particular type of case (and even some minor criminal cases) that the judge decide the facts, see CONN. GEN. STAT. § 54-82b(a) (West 2001); N.J. STAT ANN. § 2C:1-4(b) (West 2002), and judges routinely decide facts related to the admissibility of evidence, see FED. R. EVID. 104(a). But in the vast majority of cases that actually go to trial, it is a jury who decides the facts.

preponderance of the evidence standard, which requires the plaintiff to prove that the event is more likely than not; in other words, to convince the jury that it is above .5 certainty that the event occurred.[19] In criminal cases, the standard of proof is generally the beyond a reasonable doubt standard, a standard that is generally understood to require a much higher level of certainty; indeed, a level of certainty approaching (but not equaling) 1.0.[20]

---

[19] 2 MCCORMICK ON EVIDENCE, *supra* note 4, at 514. I adopt the statisticians' approach of referring to probability as a number between 0.0 and 1.0, rather than as a percentage between 0 and 100.

[20] There are alternative ways of conceiving the standard of proof. Several commentators have suggested that trials are about stories and that the standard of proof is a measure of the certainty that the jurors must have in the parties' stories. *See* LANCE BENNETT & MARTHA S. FELDMAN, RECONSTRUCTING REALITY IN THE COURTROOM: JUSTICE AND JUDGMENT IN AMERICAN CULTURE (1981); ROBERT P. BURNS, A THEORY OF THE TRIAL 121 (1999); Nancy Pennington & Reid Hastie, *The Story Model for Juror Decision Making, in* INSIDE THE JUROR: THE PSYCHOLOGY OF JUROR DECISION MAKING 192 (Reid Hastie ed., 1993); *see also* D. Michael Risinger, *John Henry Wigmore, Johnny Lynn Old Chief, and "Legitimate Moral Force": Keeping the Courtroom Safe for Heartstrings and Gore*, 49 HASTINGS L.J. 403, 440-42 (1998). The story models suggest that the standard of proof is a measure of the fitness between the parties' stories and the verdict categories, or a measure of the coherence and completeness of the stories.

Applying the story model to criminal cases has been problematic. Ronald Allen, for instance, first suggested that the model applied in civil cases, but not in criminal ones. Ronald J. Allen, *A Reconceptualization of Civil Trials*, 66 B.U. L. REV. 401, 426-28, 434-37 (1986) [hereinafter Allen, *Reconceptualization*]. More recently, however, he and Brian Leiter have stated that the story model, which they refer to as "the relative plausibility theory," requires "the prosecution [to] provide a plausible account of guilt and show that there is no plausible account of innocence." Ronald J. Allen & Brian Leiter, *Naturalized Epistemology and the Law of Evidence*, 87 VA. L. REV. 1491, 1528 (2001). As Dale Nance has noted, however, this does not appear to be a test of *relative* plausibility, in the sense that we compare the plausibility of the government's story to that of the defendant. *See* Dale A. Nance, *Naturalized Epistemology and the Critique of Evidence Theory*, 87 VA. L. REV. 1551, 1593 (2001) [hereinafter Nance, *Naturalized Epistemology*]. Nance suggests that perhaps the relative plausibility theory can be interpreted here as requiring the jurors to believe that the government's story is greater than 95% probable, given the defendant's story, as well as any other possible stories. *Id.* at 1594. If this is the proper interpretation of the story theory in the context of criminal cases, then that theory is compatible with the theory set forth in this Article.

Neil Cohen has suggested another alternative: that burdens of persuasion do not just represent the decision maker's judgment as to the probability that a particular event occurred, but also the degree of confidence that the decision maker has in that probability judgment. Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. REV. 385, 420 (1985). While Cohen's discussion focuses on the preponderance standard, his analysis can be applied to the beyond a reasonable doubt standard. *Id.* at 420-21. As a result, the reasonable doubt standard might differ from the preponderance standard not because it requires a greater point estimate, but because it requires a higher degree of confidence. *Id.* Cohen's theory, while normatively appealing, fails to describe how most legal academics and practitioners presently perceive the burden of persuasion, and lies outside the scope of this Article.

Because complete certainty is unobtainable, errors are certain to occur: there will be both errors that favor the government and errors that favor the defendant. The standard of proof determines, in part, how these errors will be allocated. For instance, it is imaginable that in criminal cases we would have a standard of proof that only let the jury convict when it was "completely certain" of the defendant's guilt.[21] Under such a standard of proof, all errors would generally fall upon the government. Few, if any, innocent persons would be convicted because the jury would "know" that the defendant was guilty, but many guilty persons would go free because the jury would not be completely certain that the defendant was guilty.

The adoption of the reasonable doubt standard is a rejection of that "complete certainty" approach and the notion that the government should bear all of the errors. According to the traditional view, the adoption of this standard is mandated by the desire to ensure that there are far more erroneous acquittals than erroneous convictions. The logic is that as the standard of proof for the government goes up, the number of erroneous acquittals proportional to erroneous convictions should also go up.

This relationship between the standard of proof and the allocation of erroneous verdicts, however, is more complicated than it first appears. This Part begins by setting forth a model of the relationship between the standard of proof and erroneous verdicts. It concludes that the precise allocation of erroneous verdicts is dependent on a number of factors, including the prior allocation of both actually guilty and actually innocent defendants at trial and the distribution of evidence against those defendants. Part I.B then formalizes the traditional understanding. In its most common form — that we require proof beyond a reasonable doubt because it is better that one innocent man go free than ten guilty men be punished — the traditional view turns out to be too vague to support any particular standard of proof. Several scholars, however,

---

Finally, some theorists reject the idea that probability has anything to do with burdens of proof. *See, e.g.*, L. JONATHAN COHEN, THE PROBABLE AND THE PROVABLE 56, 63, 248 (1977); Lea Brilmayer & Lewis Kornhauser, *Quantitative Methods and Legal Decisions*, 46 U. CHI. L. REV. 116 (1978). Again, while these theorists' ideas are interesting, they again do not reflect the way the standard of proof is normally applied. *See, e.g.*, Stephen E. Fienberg, *Misunderstanding, Beyond a Reasonable Doubt*, 66 B.U. L. REV. 651, 652-55 (1986) (explaining "weight of evidence" argument).

[21] Such a standard would presumably require subjective, as opposed to objective, certainty. Under such a subjective standard, a few errors might still occur, because the jury's subjective belief of complete certainty might be (objectively) wrong. For a discussion of objective versus subjective views of the standard of proof, *see infra* note 29.

have used decision theory, and in particular the expected utility model, to attempt to show that proof beyond a reasonable doubt can be justified using the relative utilities of erroneous acquittals and convictions. Finally, Part I.C is devoted to refining this model to account for accurate verdicts.

A. Standards of Proof, Errors, and Accuracy: An Introduction

An example may help illustrate the interaction of the standard of proof, the allocation of defendants, and the distribution of evidence in determining how many and what kinds of errors will result from litigation. Suppose there is a group of twenty defendants, ten of whom are actually guilty and ten of whom are actually innocent. Then assume that the evidence against the twenty defendants is evenly distributed, such that the evidence would leave a rational evaluator[22] somewhere between .05 and .95 certain of each of the defendants' guilt.[23] Also

---

[22] Of course, actual jurors and judges are not purely rational evaluators of the evidence. The use of actual decision makers presents another source of error, *see infra* text accompanying notes 28-29, but one that can be controlled by discussing the reasonable doubt standard as subjective, not objective in nature, in line with some current understanding. *See infra* note 29; *see also* Reid Hastie, *Algebraic Models of Decision Processes*, *in* INSIDE THE JUROR, *supra* note 20, at 84, 96-97. In addition, juries are made up of multiple jurors, each of whom is an independent evaluator of the facts in the case. *See* Bernard Grofman, *Mathematical Models of Juror and Jury Decision-Making*, *in* THE TRIAL PROCESS 305, 309 n.3 (Bruce Dennis Sales ed., 1981). For purposes of this Article, I am focusing on the role of reasonable doubt on the decision of the individual rather than the group. It is conceivable that the *jury*, as a group of people applying reasonable doubt, will impose a higher level of certainty than I describe here. For now, it is sufficient to note that even if this is true, *judges* applying reasonable doubt as triers of fact sit alone and would not be subject to this dynamic.

[23] One possible objection to this model is the conjunction or multiplication effect. *See* Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1512 (1999) [hereinafter Posner, *Economic Approach to the Law of Evidence*]; Lola Lopes, *Two Conceptions of the Juror*, *in* INSIDE THE JUROR, *supra* note 20, 255, 259-60; Allen, *Reconceptualization*, *supra* note 20, at 405-08; Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357, 1385-86 (1985) [hereinafter Nesson, *The Evidence or the Event?*]; Alan D. Cullison, *The Model of Rules & the Logic of Decision*, *in* MODELING THE CRIMINAL JUSTICE SYSTEM 225, 239-40 (Stuart S. Nagel ed., 1977) [hereinafter Cullison, *Model of Rules*]. The reasonable doubt standard is often understood to apply separately to each element of the crime charged. If the crime charged has four elements, then the jury would only have to have .66 certainty: .9 x .9 x .9 x.9 = .6561. Presumably, the same rule would apply to the preponderance of the evidence rule, which would mean that the plaintiff's case, where there is a four element civil claim, would only have to have .06 certainty: .5 x .5 x .5 x .5 = .0625. *See* Posner, *Economic Approach to the Law of Evidence*, *supra*, at 1512; Allen, *Reconceptualization*, *supra* note 20, at 405-07. Such results seem at odds with our usual understanding of these standards of proof.

assume that the criminal justice system (and in particular its mechanisms for acquiring and presenting evidence at the trial) is ineffective in sorting defendants. That is, assume that all twenty are charged with the crime and that the evidence against the twenty defendants is distributed such that there is no difference in distribution between the actually guilty and the actually innocent. The resulting allocation would resemble something like Table 1.[24]

Table 1

Certainty of Guilt	Innocent Defendants	Guilty Defendants
.05	1	1
.15	1	1
.25	1	1
.35	1	1
.45	1	1
.55	1	1
.65	1	1
.75	1	1
.85	1	1
.95	1	1

However, as a practical matter, it is not clear that the jury's decision making as to a particular element is ever "independent" so that the product rule fully applies, and often it will be conceded that one or more of the elements has been established. *See, e.g.*, Alan D. Cullison, *Probability Analysis of Judicial Fact-Finding: Preliminary Outline of the Subjective Approach*, 1969 U. TOL. L. REV. 538, 580 & 583 [hereinafter Cullison, *Probability Analysis*]; *see also* Cullison, *Model of Rules*, *supra*, at 240 (stating that inherent vagueness of reasonable doubt standard eliminates this problem); Carl. G. Wagner, *Book Review*, 1979 Duke L.J. 1071, 1076 (reviewing COHEN, *supra* note 20). *But see* Allen, *Reconceptualization*, *supra* note 20, at 409-15 (criticizing this assumption). The controversy over the conjunction effect rages on unabated. *Compare* Allen & Leiter, *supra* note 20, at 1503-05 (using existence of effect to critique expected utility theory), *with* Nance, *Naturalized Epistemology*, *supra* note 20, at 1565-75 (analyzing and critiquing Allen and Leiter). Courts and commentators, discussing both rules, seem to be requiring a high level of overall certainty for reasonable doubt, and at least .51 overall certainty for preponderance of the evidence. In any event, the multiplication effect may make an incorrect assumption: that jurors really apply it to all elements, rather than just the claim. *See* Nance, *Naturalized Epistemology*, *supra* note 20, at 1571-72. As we will see in Part II, the evidence suggests that jurors routinely ignore instructions, and there is little reason to believe that jurors are cognizant of this particular issue.

[24] There are a number of other assumptions built into this table that I will discuss below.

Now, if the standard of proof is at .5 (preponderance of the evidence),[25] the total number of defendants convicted will be ten: five guilty defendants and five innocent defendants. Similarly, ten defendants will be acquitted: five guilty defendants and five innocent defendants. There are also ten errors: five defendants who were in fact innocent have been found guilty and five defendants who were in fact guilty have been acquitted. But there are also ten cases correctly decided: five guilty defendants have been convicted and five innocent defendants have been acquitted. Note that the number of erroneous convictions and erroneous acquittals is the same.

If a reasonable doubt standard (.9) is used instead, there are a number of effects: the numbers of convictions and acquittals change, as does the allocation of errors between guilty and innocent defendants.[26] But interestingly, the number of errors will remain constant. Under the reasonable doubt standard, only two defendants are convicted: one guilty defendant and one innocent defendant. At the same time, now eighteen defendants are acquitted: nine innocent defendants are correctly acquitted and nine guilty defendants are erroneously acquitted. Furthermore, using the reasonable doubt standard, as compared to the preponderance standard, moves the number of erroneous acquittals from five to nine, and the number of erroneous convictions is lowered from four to one. So, instead of having an equal number of erroneous convictions and acquittals, there are now nine erroneous acquittals for each erroneous conviction. The overall number of errors, however, *has stayed constant*. This is because the move from the preponderance of the evidence standard to the reasonable doubt standard has traded four erroneous convictions for four erroneous acquittals. As a result, guilty defendants are nine times more likely to benefit from an error than an innocent defendant is likely to be harmed by an error.

So when we consider the effects of changes in the standard of proof on both guilty and innocent defendants, an increase in that standard can have a number of effects: acquittals (both accurate and inaccurate) will increase, while convictions (both accurate and inaccurate) will decrease. Heightening the standard of proof will lower the number of erroneous convictions (and thereby increase the number of accurate acquittals), but only at the "cost" of increasing the number of erroneous acquittals (and thereby decreasing the number of accurate convictions). Furthermore,

---

[25] 2 MCCORMICK ON EVIDENCE, *supra* note 4, at 422.

[26] Some would put the reasonable doubt standard even higher, say at .95. For present purposes, the exact percentage does not matter.

there is no guarantee that the *overall* number of errors will be affected by a change in the standard of proof. In our example, there is no effect on this number. Thus, if raising the standard of proof is going to be justified, it may have to be the result of preferring a different distribution of errors, rather than a desire to minimize the number of errors.[27]

The figures in Table 1 contain a number of assumptions. One assumption is that errors result from the amount of evidence presented. But mistakes can come from another factor: the decision maker could improperly evaluate the evidence. For instance, although there is one guilty defendant for whom the evidence rationally would result in an (accurate) conviction, the decision maker might erroneously evaluate the evidence as creating only .89 certainty, and therefore acquit.[28] Thus, there really are two sources of potential error: the evidence presented and the decision makers' evaluation of it.[29]

---

[27] *Cf.* Neil Orloff & Jery Stedinger, *A Framework for Evaluating the Preponderance-of-the-Evidence Standard*, 131 U. PA. L. REV. 1159, 1159-61 (1983) (noting that merely minimizing number of errors may not be preferable where we do not know either size of error or distribution of errors between plaintiffs and defendants).

[28] Or vice versa, the decision maker might erroneously evaluate the evidence against another defendant as creating more than (rather than less than) .90 certainty, and thereby erroneously convict.

[29] *See* ZUCKERMAN, *supra* note 11, at 122-23 (describing these as mistakes of reasoning and mistakes of fact). Zuckerman contends that "mistakes of reasoning," that is mistakes by the decision maker in evaluating the evidence as an objective matter, could be eliminated, at least in principle. *Id.* at 122. Given the limits in human reasoning and bias, however, it seems unlikely that mistakes of reasoning can be eliminated (as a practical matter) any more than mistakes of fact (i.e., mistakes that arise because of the evidence generated).

For present purposes, these two errors can be analyzed as the same error by recognizing that the reasonable doubt standard does not require an objective evaluation of the evidence, but rather a subjective assessment by the decision maker. Hastie, *supra* note 22, at 96-97. The standard itself is often described to jurors in terms that could be described as either subjective or objective. For instance, one commonly used federal model jury instruction states that a reasonable doubt "is a doubt which would cause a reasonable person to hesitate to act in a matter of importance in his or her personal life." 1 SAND et al., *supra* note 5, Instr. 4-2, at 4-8; *see also* 1A KEVIN O'MALLEY et al., FEDERAL JURY PRACTICE & INSTRUCTIONS: CRIMINAL § 12.10, at 168 (2000). Other instructions, however, use much more subjective language when describing the standard: the "evidence must be firm, convincing, believable and complete enough to leave no reasonable doubt of the defendant's guilt in *your* mind." 1 HOWARD G. LEVENTHAL, CHARGES TO THE JURY & REQUESTS TO CHARGE IN A CRIMINAL CASE: NEW YORK § 4:66, at 213 (1988 rev. ed.); *see also* 1 CALIFORNIA JURY INSTRUCTIONS: CRIMINAL Instr. 2.90, at 117 (1996).

Despite the confusion in the instructions, the Supreme Court has indicated that the standard should be seen as subjective. *See* Jackson v. Virginia, 443 U.S. 307, 315 (1979). Although I have assumed that the decision maker must be rational in evaluating the evidence presented, that assumption can be relaxed for now; instead, I only have to assume that the evidence against each defendant would be subjectively evaluated by the decision

More importantly, the figures in Table 1 make assumptions about what a population of defendants might look like. First, they assume that the criminal justice system has produced the same distribution of evidence against guilty and innocent defendants. Second, they assume that the number of innocent and guilty defendants subjected to trial are the same. Both of these assumptions have difficulties.

Most people would be shocked if the distribution of actual defendants who are tried looked anything like Table 1. Instead, we routinely presuppose that the criminal justice system does some accurate sifting of guilty and innocent suspects. For instance, if this was not true, it would mean that all individuals, guilty or innocent, have the same *ex ante* possibility of being punished by the system.[30] If that was so, there would be little reason not to engage in criminal behavior, given that the likelihood of punishment would be the same regardless of whether the individual committed the crime. This is why society invests resources in the investigation of crime: generally speaking, more evidence is generated against guilty individuals than innocent individuals.

In Table 1, though, the assumption was not about individuals as a whole, but about the defendants produced for trial. Although the system as a whole may work to create more evidence against guilty defendants than against innocent defendants, the distribution of the evidence against those who are charged could be the same.[31] In particular, plea bargaining, if it eliminates the defendants against whom there is the most evidence, might result in a distribution of evidence for guilty defendants that is very similar to that of innocent defendants.[32] However, we have little empirical evidence that suggests either result.

The second assumption — that the numbers of guilty and innocent defendants who go to trial are the same — is also hard to justify. Society as a whole has more innocent than guilty persons, and we routinely assume that the criminal justice system weeds out individuals in such a way that most defendants who are charged are, in fact, guilty.[33] Of

---

maker as being distributed in the manner set forth in Table 1.

[30] *See, e.g.*, Posner, *Economic Approach to the Law of Evidence*, *supra* note 23, at 1483.

[31] *See id.* at 1506-07. Michael DeKay assumes that there is a difference because he ignores the problem of plea bargaining. Michael L. DeKay, *The Difference Between Blackstone-Like Error Ratios and Probabilistic Standards of Proof*, 21 LAW & SOC. INQUIRY 95, 101 (1996); *see also* Grofman, *supra* note 22, at 306 n.1 (implicitly making similar assumption).

[32] Plea bargaining presents a problem for any model of reasonable doubt. Somewhat arbitrarily, I have chosen to ignore this effect.

[33] Posner certainly makes this assumption. Posner, *Economic Approach to the Law of Evidence*, *supra* note 23, at 1506, 1543. Not all commentators agree. *See* Daniel Givelber,

course, it is possible that the plea bargaining process eliminates more guilty defendants than innocent ones (either because of the strength of the evidence or because innocent defendants are less likely to agree to a plea, knowing that they are innocent).[34] Nonetheless, again there is little empirical evidence on this issue. Thus, let us assume instead for a moment that there are more guilty defendants than innocent defendants who go to trial; for now, we will say four times as many.[35] Staying with our relaxation of the equivalent distribution of evidence assumption, one then might obtain a distribution such as that set forth in Table 2.

Table 2

Certainty of Guilt	Innocent Defendants	Guilty Defendants
.05	3	4
.15	3	4
.25	3	4
.35	2	8
.45	2	8
.55	2	8
.65	2	8
.75	1	12
.85	1	12
.95	1	12

Meaningless Acquittals, Meaningful Convictions: Do We Reliably Acquit the Innocent?, 49 RUTGERS L. REV. 1317, 1341 (1997).

[34] It is also plausible that guilty defendants may be more likely to go to trial because they are less risk adverse. *See* Robert E. Scott & William J. Stuntz, *Plea Bargaining as Contract*, 101 YALE L.J. 1909, 1943 (1992).

[35] The precise ratio of guilty to innocent defendants who go to trial is not important here. The conviction rate in federal court at trial is about 80%. *See* U.S. DEPARTMENT OF JUSTICE, BUREAU OF JUSTICE STATISTICS 1998 SOURCEBOOK OF CRIMINAL JUSTICE STATISTICS 410-11 (Kathleen Maguire & Ann L. Pastore eds., 1999) (3,629 convictions and 1,081 acquittals); U.S. DEPARTMENT OF JUSTICE, BUREAU OF JUSTICE STATISTICS 1997 SOURCEBOOK OF CRIMINAL JUSTICE STATISTICS 400-01 (Kathleen Maguire & Ann L. Pastore eds., 1998) (3,730 convictions and 893 acquittals); U.S. DEPARTMENT OF JUSTICE, BUREAU OF JUSTICE STATISTICS 1996 SOURCEBOOK OF CRIMINAL JUSTICE STATISTICS 450-51 (Kathleen Maguire & Ann L. Pastore eds., 1997) (4,074 convictions and 902 acquittals). Kalven and Zeisel found a lower rate of conviction, but still one that was much greater than the rate of acquittal: 64.2% versus 30.3%. KALVEN & ZEISEL, *supra* note 7, at 56. While conviction rates and factual guilt may not be precise correlates, there is (hopefully) some correlation, which does suggest that far more guilty defendants than innocent defendants are tried. Of course, we cannot be sure that this is true.

The distribution includes 100 defendants. Now, the effect of raising the standard of proof is to alter both the ratio of errors and the total number of errors. Raising the standard of proof actually *increases* the number of errors from 35 (28 erroneous acquittals and 7 erroneous convictions) to 69 (68 erroneous acquittals and 1 erroneous conviction), nearly doubling the amount of inaccuracy. The reason for this is that a shift in the standard of proof trades more erroneous acquittals for fewer erroneous convictions. When there are more guilty defendants around to be acquitted erroneously, the shift upwards will trade a far larger number of erroneous acquittals for a smaller number of erroneous convictions.

At the same time, there is also a profound change in the ratio of erroneous acquittals to erroneous convictions ("R").[36] When the standard of proof is preponderance of the evidence, there are four erroneous acquittals for each erroneous conviction (R = 4), but when the burden is reasonable doubt, there are now sixty-eight erroneous acquittals for each erroneous conviction (R = 68). The figures in Table 1 also show a change in the ratio of errors when the standard of proof is altered. Under the preponderance standard, there is one erroneous acquittal for each erroneous conviction (R = 1), whereas under the reasonable doubt standard there is one erroneous conviction for every nine erroneous acquittals (R = 9). What is different between Tables 1 and 2 is the *magnitude* in the change of the ratio: for Table 1, the magnitude is 9 and for Table 2 it is 17. In other words, the relative numbers of guilty and innocent defendants who go to trial affects the absolute ratio of errors for various standards of proof, but it does not affect the relative ratio of errors between standards of proof. The relative ratio of errors is instead simply a function of the distribution of evidence against innocent versus guilty defendants.

This does not mean that the probabilities of acquittal or conviction for a particular innocent or guilty defendant are changed by an alteration in the relative population sizes of innocent and guilty defendants. To the contrary, the odds of conviction or acquittal under a particular standard of proof remain the same, regardless of the relative population sizes. Instead, what the relative population size alters is the probability of *all* defendants being convicted under a particular standard and the resulting number of errors. The lesson here is that, if we assume that more guilty defendants are tried than innocent defendants, then

---

[36] If the number of erroneous convictions is defined as N_{ci}, the number of erroneous acquittals as N_{ag}, then $R = N_{ag} / N_{ci}$.

increasing the standard of proof will lead to a much larger increase in the total number of errors and a much smaller decrease in the overall odds of conviction.

Changes in the standard of proof at trial do affect the outcomes of trials, although not always in completely predictable ways. Increasing the standard of proof leads to fewer convictions and more acquittals, which also means that the number of erroneous acquittals, relative to the number of erroneous convictions, will increase. The exact effect on the ratio of erroneous acquittals and convictions, however, is complicated, and the effect on the overall *number* of errors is also more difficult to determine, because it depends on the real world distribution of guilty and innocent defendants. Altering our assumptions about both the relative numbers of guilty and innocent defendants who are tried, as well as the relative strength of the evidence against these two groups, will affect how changes in the standard of proof alter both the number of errors and the ratio of errors. Furthermore, the relative number of defendants will also affect the overall odds of conviction.

The available data suggests that, in general, the evidence against those who actually go to trial is relatively high, because the vast majority of criminal trials do result in convictions.[37] Assuming then that the reasonable doubt standard does require a level of certainty somewhere well above .5, this means that in the typical criminal trial the average amount of evidence against the defendant will also tend to be well above .5. There is also some reason to believe that innocent defendants, once charged, are more likely to go to trial than guilty defendants, although there is also support for the opposite view.[38] Furthermore, we do not know the relative numbers of actually guilty and actually innocent defendants who are charged, so it is impossible to know with any degree of certainty the relative numbers of guilty and innocent defendants who in fact go to trial and are therefore subjected to a decision made under the reasonable doubt standard.[39] As a result, we have no way of accurately predicting, at present, the actual ratios of inaccurate and accurate verdicts.

---

[37] *See supra* note 35.

[38] *See supra* note 34 and accompanying text; Scott & Stuntz, *supra* note 34, at 1943.

[39] *See also* Allen, *Restoration*, *supra* note 11, at 47 n.65 (noting that "[w]ithout knowing the distribution of guilt probabilities of factually innocent and guilty defendants, we cannot know the actual effect of choosing one standard of proof over another").

B. The Traditional View and the Decision Theory Model

Just as the workings of standards of proof are more complicated than they might initially seem, so too is the traditional explanation of the reasonable doubt standard. The traditional theory has its roots in 15th Century explanations of the criminal justice system.[40] One prominent example is William Blackstone's maxim that it is far better that ten guilty men be acquitted than one innocent man be convicted, but the thinking has its genesis in John Fortescue.[41] As stated above, the traditional theory posits that the reasonable doubt standard alters the balance of these errors so as to ensure that there are more erroneous acquittals than erroneous convictions.[42] This result is viewed as enhancing social welfare because the costs of erroneous convictions are seen as heavily outweighing the costs from erroneous acquittals.[43] This proposition can be expressed by the following equation:

$$U_{ag}(N_{ag}) \geq U_{ci}(N_{ci}),^{44} \qquad (1)$$

where U_{ag} is the utility of acquitting the guilty (an erroneous acquittal) and U_{ci} is the utility of convicting the innocent (an erroneous conviction).[45] In practice, decisions to convict the innocent and acquit the guilty really create disutility. The equation, therefore, formalizes the logic that the standard of proof should be set so that the disutility from erroneous acquittals is less than or equal to the disutility from erroneous convictions.

As Professor Michael DeKay has noted, however, there are two different ways to go about viewing the relationship between these utility

---

[40] *See* Hon. John Wilder May, *Some Rules of Evidence: Reasonable Doubt in Civil and Criminal Cases*, 10 AM. L. REV. 642, 653-56 (1876) (quoting 4 WILLIAM BLACKSTONE, COMMENTARIES *558) (attributing rise of reasonable doubt standard to application of this maxim).

[41] For a history of the rise of this maxim, *see* Risinger, *supra* note 20, at 442-43 n.98; *see also* DeKay, *supra* note 31, at 95.

[42] *See* May, *supra* note 40, at 656-64 (noting and criticizing the rationale); *see also* Newman, *supra* note 9, at 980-81. On Type I and Type II errors, *see* Cullison, *Probability Analysis*, *supra* note 23, at 566; Posner, *Economic Approach to the Law of Evidence*, *supra* note 23, at 1504.

[43] *See* POSNER, ECONOMIC ANALYSIS, *supra* note 13, at 604-05.

[44] N_{ag} represents the number of erroneous acquittals and N_{ci} represents the number of erroneous convictions. U_{ag} and U_{ci} will both be negative numbers, because I assume that they actually represent disutilities. Thus, this equation stands for the proposition that the costs of convicting N_{ci} persons is greater than the costs of acquitting N_{ag} guilty persons. But as Michael DeKay has stated, it is simpler, and generally no less accurate, to represent the equation as $U_{ag}(N_{ag}) = U_{ci}(N_{ci})$. DeKay, *supra* note 31, at 104.

[45] *See, e.g., id.*

functions. The first approach is to view them, as Blackstone's maxim would seem to indicate, simply as an expression of the ratios of errors (the R value in the examples in Part I.A). Under this interpretation, what we seek is an optimal ratio of erroneous acquittals and convictions. For instance, under Blackstone's maxim, we would want the following relationship:

$$U_{ag}(10) = U_{ci}(1).^{46}$$

It is by no means clear that this is the proper ratio of erroneous acquittals to convictions. Commentators have suggested many alternatives, such as 5:1, 20:1, or even 1000:1.[47] Even if agreement is possible on the correct ratio, however, it is not easy to discover the proper point at which to set the standard of proof. As I explained in Part I.A, the ratio of erroneous acquittals to erroneous convictions is dependent on two variables: the relative distribution of evidence against innocent and guilty defendants and the relative numbers of innocent and guilty defendants who go to trial.[48] For instance, if we wanted a standard of proof that would give a ratio of nine erroneous acquittals to one erroneous conviction (i.e., R = 9), the figures in Table 1 would require that the standard of proof be set at .90. In Table 2, a standard of proof between .60 and .70 would be required.[49] Thus, the optimal standard of proof, under a ratio of errors approach, is dependent upon the distribution of evidence and the population of guilty and innocent defendants. Without more data about these two factors, it is impossible to make any predictions about the proper standard of proof.[50]

---

[46] The Supreme Court, on at least one occasion, has characterized this ratio (albeit in a slightly different context) as "not an unreasonable assumption." Ballew v. Georgia, 435 U.S. 223, 234 (1978).

[47] ZUCKERMAN, *supra* note 11, at 126; Terry Connolly, *Decision Theory, Reasonable Doubt, and the Utility of Erroneous Acquittals*, 11 LAW & HUM. BEHAV. 101, 104 (1987); DeKay, *supra* note 31, at 104.

[48] *See supra* text accompanying notes 22-36. DeKay points out that the relative amounts of evidence ("the prior odds of guilt") matter, DeKay, *supra* note 31, at 110, and he acknowledges that the effects of a decision criterion will depend on the prior odds of guilt and the effects of accuracy within the criminal justice system. *See id.* at 121-25; *see also id.* at 118 & n.59 (citing Allen, *Restoration*, *supra* note 11, at 47 n.65). DeKay talks about the distributions of guilty and innocent defendants but assumes that both groups fit into bell-shaped curves, an assumption that this Article does not make. *Id.* at 101, 110; *see also* Grofman, *supra* note 22, at 305, 311-12.

[49] At .70, R = 14.6, because there are 44 erroneous acquittals and 3 erroneous convictions. At .60, R = 7.2, because there are 36 erroneous acquittals and 5 erroneous convictions.

[50] *See* DeKay, *supra* note 31, at 119-26. For a recent discussion of these issues in connection with the preponderance of the evidence standard, *see* D.H. Kaye, *The Error of*

One way to avoid these problems is to adopt an alternative approach to Blackstone's maxim and interpret the maxim as a ratio not of errors, but rather of *utilities* of the errors.[51] Professors John Kaplan and Alan Cullison, working independently, set forth what this Article refers to as the decision theory model.[52] Drawing on decision theory and the expected utility model, the decision theory model might provide support for the idea that the criminal standard of proof ought to be and is quite high.[53] This model begins with the assumptions that: (a) the probability of guilt may be expressed as a function of utilities associated with acquittals and convictions and (b) the standard of proof should be set at the point where the desirability of convicting would either be equal to or above the desirability of acquitting.[54] Working from these two assumptions, the model then demonstrates that the standard of proof (*SP*) in any case, civil or criminal, is a function of the utility of convicting an innocent person (U_{ci}) and the utility of acquitting a guilty person

---

Equal Error Rates, 1 LAW, PROBABILITY & RISK 3 (2002).

[51] DeKay, *supra* note 31, at 106-09.

[52] *See* John Kaplan, *Decision Theory and Reasonable Doubt*, *in* COMMUNICATION SCIENCES AND THE LAW: REFLECTIONS FROM THE JURIMETRICS CONFERENCE (Layman E. Allen & Mary E. Caldwell eds., 1965) [hereinafter Kaplan, *Decision Theory and Reasonable Doubt*]; John Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065 (1968) [hereinafter Kaplan, *Decision Theory and the Factfinding Process*]; Cullison, *Probability Analysis*, *supra* note 23; *see also* DeKay, *supra* note 31, at 97 & n.9. Although there is some variation on how they arrive at their models, the resulting models for both Kaplan and Cullison are virtually identical, and I will treat them as such in this Article. *See also* Dale A. Nance, *Evidential Completeness and the Burden of Proof*, 49 HASTINGS L.J. 621, 623 (1998) [hereinafter Nance, *Evidential Completeness*].

[53] It is unclear to what extent, if any, Kaplan and Cullison saw their models as *normative* explanations of the reasonable doubt standard. *See* Patricia G. Milanich, *Decision Theory and Standards of Proof*, 5 LAW & HUM. BEHAV. 87, 93 (1981) (noting that it is unclear what Kaplan is trying to do). *Compare* Cullison, *Model of Rules*, *supra* note 23, at 225, 237-39 (stating that this theory is normative), *and* Cullison, *Probability Analysis*, *supra* note 23, at 576 (seeming to suggest it may have normative uses), *with* Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1032 (1977) (stating that this theory is not normative).

[54] *See* Cullison, *Probability Analysis*, *supra* note 23, at 565 (equal); Kaplan, *Decision Theory and the Factfinding Process*, *supra* note 52, at 1071 (greater than). DeKay adopts Kaplan's approach. *See* DeKay, *supra* note 31, at 110.

(U_{ag}),[55] such that:

$$SP = \frac{1}{\left(1 + \frac{U_{ag}}{U_{ci}}\right)}.^{56} \qquad (2)$$

In other words, the amount of certainty required by the standard of proof will go up as the costs of an erroneous conviction rise relative to the costs of an erroneous acquittal. For instance, if we assume that the costs of errors favoring the plaintiff and defendant are equal (in other words, that $U_{ci} = U_{ag}$), the model will predict that the level of certainty required by the standard of proof is .50, which seems to reflect what is required by the preponderance of the evidence standard.[57] If instead we assume that the disutility of an erroneous conviction is ten times the disutility of an erroneous acquittal (for example, say $U_{ci} = -10$ and $U_{ag} = -1$), then the standard of proof rises to .91, which would be quite close to the .9 standard we have discussed.[58]

C. Refining the Model: Accurate Verdicts

One immediate problem for Kaplan and Cullison's model is that it ignores the effects of the utilities associated with the accurate verdicts: the conviction of guilty persons and the acquittals of innocent persons,

---

[55] In a civil case, instead of discussing "convicting an innocent person," U_{ci} would refer to holding a person liable even though not legally responsible, and U_{ag} would refer to not holding a person liable even though the person was legally responsible. *See* Cullison, *Probability Analysis*, *supra* note 23, at 568.

[56] *See Kaplan, Decision Theory and the Factfinding Process*, *supra* note 52, at 1072. For another version of this model based on the ratio of U_{ci} to U_{ag}, *see* Grofman, *supra* note 22, at 313-14. *See also* Stuart Nagel et al., *Decision Theory and Juror Decision-Making*, *in* THE TRIAL PROCESS, *supra* note 20, at 353, 358-59 (making similar simplifying assumption).

This formula is based on the further assumption that the burden of proof should be set at the stage where the expected utility (EU) of convicting (EU(C)) is equal to the expected utility of acquitting (EU(A)). The expected utility is, in turn, a function of the utility and the probability of guilt, such that $EU(x) = p * u$, where p is the probability of that outcome and u is the utility of such an outcome. Because the standard of proof is set at the probability where the expected utility of the choices of acquitting and convicting are the same, Kaplan assumes (implicitly) that $EU(A) = SP * U_{ag}$ and $EU(C) = (1\text{-}SP) * U_{ci}$. *See* Kaplan, *Decision Theory and the Factfinding Process*, *supra* note 52, at 1071-72. Then, solving for SP, we get $SP * U_{ag} = U_{ci} - (SP * U_{ci})$, such that $(SP * U_{ag}) + (SP * U_{ci}) = U_{ci}$, and $SP(U_{ag} + U_{ci}) = U_{ci}$, finally leading to $SP = U_{ci} / (U_{ag} + U_{ci}) = 1/(1 + U_{ag}/U_{ci})$.

[57] *See* Kaplan, *Decision Theory and the Factfinding Process*, *supra* note 52, at 1072.

[58] *See* Milanich, *supra* note 53, at 88-89; Nagel et al., *supra* note 56, at 357. Again, the assumption that $U_{ci} = 10 * U_{ag}$ is a common one. *See supra* note 46 and accompanying text.

an insight first made by Professor Laurence Tribe.[59] Accounting for the utility of convicting the guilty (U_{cg}) and the utility of acquitting the innocent (U_{ai}) results in a far more complex equation for arriving at the standard of proof (*SP*):

$$SP = \frac{1}{1 + \left(\frac{(U_{cg} - U_{ag})}{(U_{ai} - U_{ci})} \right)} \quad (3)$$ [60]

In other words, the standard of proof should be set by weighing the relative utilities (or disutilities) of both accurate and erroneous convictions and both accurate and erroneous acquittals.

This version of the equation makes it far harder to reach conclusions about the value of *SP*. For the value of *SP* to be high (that is, equal to or greater than .90), it is no longer enough that the disutility of an erroneous conviction be much greater than the disutility of an erroneous acquittal. Using Equation 3, the value of the denominator of our fraction ($U_{ai} - U_{ci}$) has to be several times greater than the numerator ($U_{cg} - U_{ag}$) to result in a value of .90 — in fact, the denominator must be ten times greater than the numerator. This means that the value of decisions to convict the innocent (U_{ci}) not only must be much lower than the value of decisions to acquit the guilty (U_{ag}), but that the values of convicting the guilty (U_{cg}) and acquitting the innocent (U_{ai}) are not set in a way that will offset the difference between U_{ci} and U_{ag}. Finally, what matters is not the actual utilities (or disutilities) associated with each utility function, but

---

[59] *See* Tribe, *Trial By Mathematics*, *supra* note 13, at 1378-81; *see also* STUART S. NAGEL & MARIAN G. NEEF, DECISION THEORY AND THE LEGAL PROCESS 189-90 (1979); Nagel et al., *supra* note 56, at 358; *see also* Martin F. Kaplan, *Cognitive Processes in the Individual Juror*, *in* THE PSYCHOLOGY OF THE COURTROOM, *supra* note 6, at 197, 215-16 [hereinafter Kaplan, *Cognitive Processes*] (explaining Tribe's logic). Cullison is aware of the issue but skims over it by redefining $U_{ci} + U_{ai}$ as "Type I error" and $U_{ag} + U_{cg}$ as "Type II error" without acknowledging directly that they also contain utilities associated with accuracy. *See* Cullison, *Probability Analysis*, *supra* note 23, at 565-66. However, he later notes that jurors may alter the values of all the variables, which will result in changes to the standard of proof. *See id.* at 567.

[60] *See* DeKay, *supra* note 31, at 111. This equation is derived by recognizing that EU(C) really stands for the utility of convicting a guilty person, U_{cg}, plus the disutility of convicting the innocent, U_{ci}, and EU(A) really stands for the utility of acquitting the innocent, U_{ai}, plus the disutility of acquitting the guilty, U_{ag} meaning that $EU(C) = (p * U_{cg}) + ((1 - p) * U_{ci})$, and $EU(A) = ((1 - p) * U_{ai}) + (p * U_{ag})$. Solving for *SP* as the *p* at which EU(C) = EU(A), we get $SP = 1/(1 + ((U_{cg} - U_{ag})/(U_{ai} - U_{ci})))$. It is possible to return from this equation to Equation 2 if we assume that $U_{cg} = - U_{ag}$ and that $U_{ai} = - U_{ci}$. *See* NAGEL & NEEF, *supra* note 59, at 188-89.

instead the *relative* values of these numbers with respect to each other.[61]

To show how this works, a few examples may be instructive. Tribe, for instance, suggests the following set of values for the various utilities:

$$U_{cg} = 1, U_{ai} = 2/3, U_{ag} = 1/2, U_{ci} = 0.^{62}$$

In other words, there is no utility (positive or negative) for an erroneous conviction, there is a positive utility of 1 for accurate convictions, 2/3 utility for an accurate acquittal, and 1/2 utility for an erroneous acquittal. The result of these assumptions is that $SP = .57$.[63] Of course, the values that Tribe assigns to the four utility variables may or may not be sound.[64] Indeed, one could instead set the numbers at:

$$U_{cg} = 1, U_{ai} = 0, U_{ag} = -1, U_{ci} = -10.$$

Here, we begin with the assumption that there is neither a utility nor a disutility associated with acquitting an innocent person.[65] Then, we assume that there would be a positive social utility of 1 for convicting a guilty person, therefore the (dis)utility associated with this is – 1. However, the disutility of convicting a single innocent person is much higher: it is ten times the utility created by a single conviction of a guilty person. Furthermore, this assumes that there is also a disutility

---

[61] *See* Cullison, *Probability Analysis*, *supra* note 23, at 585 n.67; DeKay, *supra* note 31, at 108; Lempert, *supra* note 53, at 1032; Nagel et al., *supra* note 59, at 357 n.2.

[62] Tribe, *Trial by Mathematics*, *supra* note 13, at 1379-80.

[63] *See id.* at 1380 (giving this number in its fractional form: 4/7). This is because $SP = 1/(1 + ((1 - .5)/(2/3 - 0)) = .57$.

[64] *See* Milanich, *supra* note 53, at 90 (referring to them as unintuitive); *see also* Connolly, *supra* note 47, at 104. Tribe arrives at these numbers by first assuming that convicting the guilty is the outcome the trier of fact should like the most and that convicting the innocent is the least preferable outcome (to the trier of fact). The value for acquitting the innocent is derived by assuming that the trier of fact would view acquitting the innocent as 2/3 the value of convicting the guilty, while acquitting the guilty would only be worth 1/2. *See* Tribe, *Trial by Mathematics*, *supra* note 13, at 1379-80.

Patricia Milanich gives her own numbers that she claims reflects our settled values: $U_{cg} = .9$, $U_{ai} = 1$, $U_{ag} = .1$, $U_{ci} = 0$. Milanich, *supra* note 53, at 91. Solving for Milanich's values, the result is $SP = .55$. Obviously, Milanich's "settled" values result in a number far lower than we would expect; indeed lower than Tribe's number. *See also* DeKay, *supra* note 31, at 116-117 (making same point); Connolly, *supra* note 47, at 104 (making same point). It is debatable whether Milanich has correctly ordered the values. For instance, Stuart Nagel and colleagues have come up with their own calculation that mirrors the legal system's ideology. Nagel and his colleagues give the values $U_{cg} = 10$, $U_{ai} = 100$, $U_{ag} = -10$ and $U_{ci} = -100$, which results in $SP = .91$. *See* Nagel et al., *supra* note 56, at 355-56. This number can also be arrived at by assuming that $U_{cg} = 0$, $U_{ai} = 0$, $U_{ag} = -1$ and $U_{ci} = -10$. *See* DeKay, *supra* note 31, at 116.

[65] This may or may not be an acceptable assumption. But even if one does not agree that $U_{ai} = 0$ in any situation, the important point is that the varying of the values of the various utility functions relative to one another will alter the resulting standard of proof.

associated with acquitting a guilty person but that this error is, as Blackstone asserts, one-tenth that of the cost of an erroneous conviction.

Working from these assumptions, the result is *SP* = .83, a number that is closer to .91 than the number Tribe generates, but still not quite there.[66] To actually reach .91, we could change our numbers so that the disutility of an erroneous acquittal is 0. If so, now *SP* = .91. Alternatively, the value of acquitting the innocent could be changed to 10, which will also result in *SP* = .91.[67]

Before we get too comfortable with these possible improvements on the model, note that it turns on the particular values assigned. If, instead, the positive utility of an accurate conviction is one-half the disutility of an erroneous conviction, keeping the other numbers the same from the original assumptions, then the value of *SP* is reduced to .63. And assigning a positive utility to an accurate acquittal has its own distorting effects.[68]

The result of all of this is that it is hard to know what value the traditional view should give to the standard of proof without some sort of idea about the underlying values of utilities of both accurate and inaccurate verdicts.[69] Depending on the values we assign, the resulting level of certainty can either be quite high or quite low. So to come to any conclusions about what the standard of proof is or ought to be under Kaplan and Cullison's model, we need to know more about the relative values of each of these functions.

---

[66] *See infra* Part II.

[67] *See supra* note 63 (setting forth Nagel's numbers).

[68] Say that instead, we have $U_{cg} = 5$, $U_{ai} = 5$, $U_{ag} = -1$ and $U_{ci} = -10$. Then the value for *SP* is .71. One could assign other values and come up with other values for *SP*. *See, e.g.*, Robert J. MacCoun, Modeling the Impact of Extralegal Bias and Defined Standards of Proof on the Decisions of Mock Jurors and Juries 22 (1984) (unpublished Ph.D. dissertation) (on file with author).

[69] *See* Itzhak Gilboa & David Schmeidler, *Case-Based Decision Theory*, 110 Q.J. ECON. 605, 625 (1995) (making this point about any use of expected utility theory). Terry Connolly has suggested a way out of this quandary by simply assuming that correct outcomes have the same maximum utility (1.0) and that convicting the innocent always has the same minimum utility (0.0). (Connolly assumes that all utility values go from 0.0 to 1.0, an acceptable assumption because, as was noted in note 48, *supra*, the only important thing is that the metric be adjusted to account for *relative* utility values.) For Connolly, the only variable then is the relative utility of U_{ag} (the utility of an erroneous acquittal). Connolly, *supra* note 47, at 109.

Even with these assumptions, however, Connolly concedes that the value for *SP* could vary from anywhere between .5 and 1.0. *Id.* at 109-10. And, as I suggest in Part III, Connolly's assumptions may be suspect. For instance, there are good reasons to believe that U_{ai} and U_{cg} will diverge in particular cases and that we cannot easily assume that either U_{ai} or U_{cg} is the maximum relative utility function.

Nonetheless, the traditional view of reasonable doubt, as modified by decision theory and the expected utility model, has the potential to explain reasonable doubt. If the appropriate relative values are assigned to the various utility functions in Equation 3, it is possible to come to a high value for *SP*. This raises two questions, one descriptive and the other normative. First, do jurors value the reasonable doubt standard in a way that is consistent with the traditional understanding? Second, if the answer to the first question is no, then how should we want jurors to value reasonable doubt? It is these questions to which this Article turns in Parts II & III.

II. THE PROBLEM OF JURORS: THE FAILURE OF THE TRADITIONAL VIEW AS A DESCRIPTIVE MODEL

Despite the superficial plausibility of the traditional view, even as supplemented by decision theory, it is simply not supported by the empirical evidence. The traditional understanding assumes that a high level of certainty is required before jurors will convict a defendant, but the data shows that this is not true. Instead, it is clear that jurors will convict at a level well below that which the traditional understanding predicts. Furthermore, the studies to date also strongly suggest that the precise level of certainty required in any criminal case will vary, depending on a number of variables.

Although this conclusion is inconsistent with the traditional understanding, it is consistent with decision theory and the expected utility model. In Part II.B, this Article sets forth a model of how we should expect jurors to reach decisions. First, it shows that even under a traditional expected utility model of juror behavior, we should expect jurors to be willing to vary the standard of proof, as the data suggests. It then incorporates the insights of behavioral economics to show that even when we drop the assumption of juror rationality, the standard of proof will still be lower than the traditional view suggests. Finally, drawing upon the social norms literature, this Article points out that the existing legal and social norms are so relatively powerless that it should not be surprising that jurors routinely vary the standard of proof.

A. Measuring Reasonable Doubt: The Empirical Data

The traditional view suggests that the reasonable doubt standard of proof requires the jurors to have a great deal of certainty in the defendant's guilt prior to convicting the defendant. The empirical evidence belies this claim. Reid Hastie has summarized over 15 different

empirical studies of the reasonable doubt standard in experimental trial settings.[70] The first, and largest, group of studies involved subjects (judges, citizens and/or students) directly stating what degree of certainty they would require for the reasonable doubt standard. The results of these surveys vary greatly: all the way from .92 certainty to .51.[71] One natural inference from the studies is that, under varying conditions, jurors will alter the standard of proof.

An alternative explanation of Hastie's meta-analysis is that the underlying studies are not consistent in their methodology, thereby resulting in variances in the data. For instance, someone could criticize many of the studies for surveying judges or students, groups that may not be representative of the actual population of jurors (the same critique could be made of using "citizens," although presumably surveys of "citizens" are more representative than surveys of judges or students). As for surveys of judges, the critique might hold, but for students it may not be valid.[72] Even if we look just at surveys of citizens, however, there is still a significant variation in the decision criteria reported in direct measurement studies: from .92 to .79.[73]

A similar objection would be that directly asking research subjects to report their decision criteria may not be the most accurate way to obtain this data; in particular, subjects may not be aware of the values that they actually use or, even if they are, they may give answers that reflect the socially proper response, rather than the level of certainty they themselves would use.[74] Other research, though, suggests that this is not an explanation for the variance that Hastie has uncovered. Another set of studies asked subjects to value the utility functions in Equation 3;

---

[70] *See* Hastie, *supra* note 22, at 101-06.

[71] *See id.* at 102 table 4.1.

[72] The evidence suggests that judges may have higher decision criteria than citizens, *see* Hastie, *supra* note 22, at 107, a result that is inconsistent with Kalven and Zeisel's suggestion that jurors are more reluctant to convict than judges. *See* KALVEN & ZEISEL, *supra* note 7, at 58. As for students, some researchers have found that there are no differences in the decisions made by groups of students and groups of citizens called for jury duty, while others have found differences. *Compare* Jonathan D. Casper & Kennette M. Benedict, *The Influence of Outcome Information & Attitudes on Juror Decision Making in Search & Seizure Cases*, *in* INSIDE THE JUROR, *supra* note 20, at 78 (finding no differences), *with* Robert J. MacCoun & Norbert L. Kerr, *Asymmetric Influence in Mock Jury Deliberation: Jurors' Bias for Leniency*, 54 J. PERSONALITY & SOC. PSYCHOL. 21, 24 (1988) (finding differences), *and* Rita James Simon & Linda Mahan, *Quantifying Burdens of Proof: A View from the Bench, the Jury and the Classroom*, 5 LAW & SOC'Y REV. 319, 322 (1971) (also finding differences).

[73] Hastie, *supra* note 22, at 102 table 4.1.

[74] *See* Stuart Nagel, *Bringing the Values of the Jurors in Line with the Law*, 63 JUDICATURE 189, 191 (1979).

from this information, the researchers calculated the standard of proof. These surveys, which are discussed in more detail in Part III.D, reported results ranging from .58 to .50.[75] While it is possible to calculate the level of certainty jurors (either individually or on average) should prefer in deciding criminal cases, using values arrived at through these surveys when we are searching for the *actual decision criteria* used by decision makers requires an additional assumption: that jurors can accurately transition from evaluating their personal (dis)utility values to the standard of proof. This assumption may not hold.[76]

A third set of studies (the "Rank-Order" studies[77]) attempts a perhaps more sophisticated approach to measuring the decision criteria actually applied by subjects. The Rank-Order studies compare the various levels of certainty that one set of subjects had in a defendant's guilt with the actual decisions as to innocence or guilt made by the same or another group of subjects. Such studies have reported results that range between those given by the other two methods: from .80 to .525. In the first of these, Rita Simon performed an experiment with students in an introductory sociology class where she had the subjects watch a 40 minute tape of a first degree homicide trial and also read newspaper clippings concerning the events.[78] She then asked half of the subjects to give their opinion as to whether the defendant was guilty. The other half were asked to describe "the probability you have in mind that the defendant . . . committed the Act."[79] The subjects did this by marking the spot between a 10 out of 10 chance and a 0 out of 10 chance (divided into

---

[75] *See infra* text accompanying notes 190-200.; *see also* Hastie, *supra* note 22, at 105 table 4.3. One arguable exception is Nagel's study (discussed *infra* at text accompanying notes 189-93) that reported results ranging from .68 to .90, but only after the use of various jury instructions. Those results, therefore, do not support the notion that there is a social norm favoring a higher standard of proof. Nagel obtained a result of .68 by giving a reasonable doubt "instruction" that gave no definition of reasonable doubt, and then achieved results of .80 by giving an instruction that required 90% certainty and .90 by giving an instruction that required a 10 to 1 Blackstone error ratio, all in the context of a rape case that resulted (in most cases) in different probability thresholds for men and women. Nagel, *supra* note 74, at 194-95.

[76] *See* Nagel, *supra* note 74, at 193 (noting that subjects were not consistent in voting to convict when subject's perception of likelihood of guilt exceeded threshold that subject should have used under expected utility theory); *see also* Nagel et al., *supra* note 56, at 375. *But see infra* notes 283-86 and accompanying text (noting that such surveys have been shown to be accurate predictors of jury behavior).

[77] *See* MacCoun, *supra* note 68, at 19.

[78] Rita James Simon, *"Beyond a Reasonable Doubt" — An Experimental Attempt at Quantification*, 6 J. APPLIED BEHAV. SCI. 203, 204 (1970).

[79] *Id.* at 204.

.5 chance intervals).[80] She then arrayed these responses along a scale.

Simon performed the experiment for two different mock defendants and two different set of newspaper clippings. She found that between 19% and 23% of the first group of subjects voted to convict (the "reference percentage"). She then looked to the data set from her second group and counted down from the position of complete certainty (10 out of 10 chance that the defendant committed the act) until her subset of subjects equaled the reference percentage. The lowest level of certainty reported by members of this subset was between .74 and .80 certainty of the defendant's guilt.[81] Simon's results thus suggest (assuming that the first and second sets of subjects were equivalent) that her subjects would require somewhere between .74 and .80 in their own subjective certainty before voting to convict the defendant.[82]

Two other researchers have also reported results suggesting that the actually applied standard of proof is lower than what we usually assume. Francis Dane tested the degree of certainty under a reasonable doubt instruction using Simon's approach.[83] Unlike Simon, however, Dane had each of his student-subjects rate both the probability of commission of a crime and state a preferred verdict (guilty or not guilty).[84] Dane found that, prior to deliberation, his subjects produced a reasonable doubt value of .725.[85] He also found that *after* deliberation, the level of certainty that his subjects had required had dropped to .525.[86] Robert J. MacCoun performed a similar study.[87] As in Dane's study,

---

[80] *Id.* at 204-05.

[81] *Id.* at 206-07. Counting from zero (i.e., from the position of no certainty) and tying the percentage to those who voted not guilty, Simon found that proof not beyond a reasonable doubt began somewhere between .70 and .72. *Id.*

[82] The subjects do not appear to have deliberated after hearing the trial. *See id.* at 204 (stating that when recording was over, students were given ballot).

[83] Francis C. Dane, *In Search of Reasonable Doubt: A Systematic Examination of Selected Quantification Approaches*, 9 LAW & HUM. BEHAV. 141 (1985). Dane's subjects read a transcript of an aggravated assault case in which the defense was self-defense. He used three different versions of the transcript, which had been pre-tested to produce conviction rates of either 75%, 50% or 25%. *Id.* at 146-47. The subjects were then asked a series of questions, including their verdict choices and the likelihood that the defendant committed the crime. *Id.* at 147. In addition, the subjects were asked their subjective confidence in their verdict, a value for the four possible verdict choices, and the minimum likelihood of guilt that the subject would require before voting for the defendant's guilt. *Id.* The subjects then deliberated in a group until they reached a unanimous verdict. After deliberation, they were then given a virtually identical questionnaire. *Id.* at 147-48.

[84] *Id.* at 143.

[85] *Id.* at 150.

[86] *Id.*

[87] MacCoun, *supra* note 68. The student-subjects were given a written transcript of an

after reading and listening to the trial, the student-subjects were given a questionnaire in which they were asked to estimate both the probability that the defendant committed the crime and the individual's preferred verdict (guilty or not guilty).[88] The subjects then deliberated in juries for up to 30 minutes, after which they filled out another questionnaire.[89] The result for MacCoun was that a reasonable doubt instruction yielded a .56 decision criterion.[90] The Rank-Order studies, therefore, appear to confirm that the applied standard of proof is lower than the traditional view assumes.

There are, of course, reasons to be wary of drawing too many conclusions from these studies, despite their apparent superiority. First, the Rank-Order studies are limited in number and deal with samples that may be too small. Second, there may be various methodological problems: (a) the subjects are almost all students,[91] (b) the studies are inconsistent on whether a subject should give both a verdict choice and a degree of certainty necessary to convict,[92] and (c) the studies used relatively short materials to educate the subjects on the trial itself, and therefore may not reflect what happens at a real world trial.

Nonetheless, the Rank-Order studies, in conjunction with the other empirical data, do strongly suggest, or at least in some cases, that subjects are willing to convict at levels of certainty lower than the traditional understanding commonly assumed. Furthermore, Dane's and MacCoun's studies both suggest that the Rank-Order method is the

---

imaginary auto theft trial — including opening statements and closing arguments, direct and cross-examinations and jury instructions — which was also played for the subjects as an audio tape recording (the subjects were split into two groups for instruction purposes: one group received a standard reasonable doubt instruction, the other received a standard preponderance of the evidence instruction). *Id.* at 45, 51-52. The evidence against the defendant was circumstantial and the defendant proffered explanations for the government's evidence. *Id.* at 46. The evidence against the defendant was sufficiently weak that the conviction rate for those subjects who received a reasonable doubt instruction was only 44.6%. *Id.* at 60, table 3.

[88] *Id.* at 53-55, 118.

[89] *Id.* at 53, 55, 124.

[90] *Id.* at 74, table 10. MacCoun's research yielded a decision criterion of .53 for a preponderance of the evidence instruction. *Id.* MacCoun is unclear, but these results appear to be based on the pre-deliberation questionnaires.

[91] *But see supra* note 72 (discussing conflict in literature over accuracy of use of students as subjects).

[92] Dane argues that this is necessary to create cross-validation. Dane, *supra* note 83, at 143; *see also* MacCoun, *supra* note 68, at 19. The difficulty is that the subject's response to one question may be affected by her response to another question, a problem that may be exacerbated by Dane's and MacCoun's methodology of asking the subjects a large number of questions related to the verdict.

most accurate of the existing methods for estimating subjects' actual decision criteria.[93] Accordingly, the empirical evidence appears to confirm that jurors will not always require a high level of certainty in deciding to convict a defendant.

Showing that the standard of proof is not just lower than the traditional understanding presumes, but also varies from case to case, requires an additional step.[94] The few Rank-Order studies give some, but certainly not conclusive support to the notion that jurors will vary the standard of proof depending upon the circumstances. MacCoun's test yielded a decision criterion of .56 and Dane (after deliberation) yielded a decision criterion of .525. Simon's tests yielded a higher level of certainty, somewhere between .74 and .80, but it appears that she did not allow the subjects the opportunity to deliberate, and Dane's test, prior to deliberation, showed a standard of proof around .725.[95]

The other studies on the standard of proof also show a great deal of variation in the level of certainty.[96] However, it is difficult to know whether this variation is the result of the particulars of the case or the result of the research methods. Additional support comes from a study by Rita Simon and Linda Mahan.[97] They surveyed judges, students, and persons called for jury duty but not actually serving on a jury. As part of the survey, the subjects were asked to state what probability of guilt (from 0 to 10) they would require for a list of offenses. The mean responses varied from 9.2 (judges' certainty for murder and jurors' certainty for manslaughter) to 7.4 (jurors' certainty for petty larceny) and the median responses varied from 9.5 (jurors' certainty for murder) to 7.5 (jurors' certainty for petty larceny). Perhaps most significant were the responses of the subjects who had actually been called for jury duty. These subjects, even more than the judges and students, showed a great deal of variation between crimes and did require greater certainty for the more serious crimes (murder and manslaughter) than they did for less serious crimes (particularly petty larceny).[98]

Other evidence also points to jurors varying the standard of proof. For instance, Nagel and Neef's work has shown that in a rape case both men

---

[93] *See* Dane, *supra* note 83, at 150; MacCoun, *supra* note 68, at 90.

[94] Nonetheless, others also assume that this occurs. *See* Power, *supra* note 9, at 49 (stating "juries will necessarily apply a range of standards to the evidence").

[95] This suggests, contrary to other data, that deliberation lowers the standard of proof.

[96] *See* Hastie, *supra* note 22, at 102 table 4.1, 105 table 4.3.

[97] *See* Simon & Mahan, *supra* note 72.

[98] *Id.* at 325, 329 & table 10.

and women would change the level of certainty they would use.[99] They also found variability in the standard of proof when different crimes were tried before different demographic groups.[100] Furthermore, empirical studies suggest that various types of evidence can affect jurors' decisions to convict or acquit. Evidence of the character of the defendant appears to increase or decrease whether jurors will vote to acquit.[101] It has also been shown repeatedly that evidence of a defendant's past conviction increases the likelihood that the jurors will vote to convict.[102] Conversely, jurors are more willing to acquit where the defendant has already been punished or where the jurors believe that the punishment is too severe, but more willing to convict where the crime is shocking.[103] These different types of information may affect conviction rates because they make it more or less likely that the defendant actually committed the crime, but these types of information may also affect conviction rates because they alter the value that jurors assign to *SP*.

What is perhaps most clear is that the studies provide *no* support for the idea that the reasonable doubt standard, as applied by jurors, is a high, fixed standard. To the contrary, the evidence strongly suggests that the standard of proof is much lower *and* that jurors routinely vary the amount of proof needed to convict.

B. Modeling Reasonable Doubt: Decision Theory and Jurors

Accepting that jurors do vary the standard of proof in criminal cases has the potential to save the decision theory version of the traditional view of reasonable doubt. As Kaplan and Cullison originally set forth their model in Equation 2, it appeared that jurors would require a high level of certainty in all criminal cases because they would place so much more negative utility on erroneous convictions than on erroneous acquittals. The need to account for accurate convictions, which results in Equation 3, upsets this conclusion; instead, as Part I.C notes, proper use

---

[99] *See* NAGEL & NEEF, *supra* note 59, at 197.

[100] *See id.* at 197-98.

[101] *See* VALERIE P. HANS & NEIL VIDMAR, JUDGING THE JURY 133-34, 143-44 (1986).

[102] *See id.* at 126.

[103] *See* Ariam Elwork et al., *The Trial: A Research Review*, *in* THE TRIAL PROCESS *supra* note 22, at 11, 12, 28; *see also* Francis C. Dane & Lawrence S. Wrightsman, *Effects of Defendants' and Victims' Characteristics on Jurors' Verdicts*, *in* THE PSYCHOLOGY OF THE COURTROOM, *supra* note 6, at 83, 97-98 (noting that defendants' remorse or suffering affects verdict outcomes). Empirical studies suggest that jurors do actually discuss the consequences of convictions during deliberations. Garold Stasser et al., *The Social Psychology of Jury Deliberation: Structure, Process & Product*, *in* THE PSYCHOLOGY OF THE COURTROOM, *supra* note 6, at 221, 232.

of decision theory seems to suggest that individuals measure the relative utilities of the various functions in such a way that the decision maker might set the standard of proof much lower. Certainly Laurence Tribe did — recall that his numbers led to a standard of proof set at .57 — and although my numbers were higher, I was able to vary the value of *SP* anywhere from .61 to .91, based on differing sets of assumptions.

This Part begins by showing that there are strong reasons to believe that jurors will not require the high level of certainty predicted by the traditional theory — largely because they do vary the actual utilities and disutilities that they bring with them to the criminal trial. Careful scrutiny of those utilities and disutilities suggests that jurors require — at least in some cases — far less certainty than we usually associate with reasonable doubt. This Part then uses insights from behavioral economics to suggest that jurors may also require less certainty than the traditional theory assumes because they might tend systematically to overvalue convictions and undervalue acquittals. Finally, this Part discusses various constraints (formal and informal) that might have the effect of increasing the amount of certainty that jurors actually require.

1. Jurors as Rational Actors

We can start with the question: why should jurors require a high level of certainty before returning a guilty verdict? One approach is to suggest that jurors simply act to maximize their own rational self-interest. This might lead jurors, in deciding upon the appropriate standard of proof, to weigh their own personal utility functions.[104] If jurors in fact do place a much greater disutility on erroneous convictions than on erroneous acquittals, this might then suggest that jurors will require a high level of certainty in criminal cases. So, returning to Equation 2, if jurors place ten times as much disutility on erroneous convictions as on erroneous acquittals, we would predict that jurors would, in practice, require a standard of proof close to .90.[105]

Of course, as we saw in Part I.C, this ignores the effects of the utilities associated with accurate convictions. In addition, it also makes a large assumption about how jurors themselves order the utilities and

---

[104] Both Kaplan and Cullison understood that the model could be used as a heuristic to describe how a particular rational juror would decide the case. *See* Cullison, *Probability Analysis*, *supra* note 23, at 564 (describing his model as "outline . . . for description and analysis of judicial fact-finding"); *see also* Milanich, *supra* note 53, at 93 (questioning whether Kaplan's goal is normative or descriptive).

[105] *See supra* text accompanying note 58.

disutilities. For instance, if I was a juror and was to order my utilities and disutilities in the manner I suggested above, my actual standard of proof would only be .83, not .90.[106]

Indeed, it seems quite likely that jurors' own personal utility functions are likely to be such that, at least in some circumstances, the applied level of certainty should vary from the high level of certainty suggested by the traditional theory. As Cullison himself states, "[a]lthough there are numerous controls on the [jurors] to keep them within some bounds, one must acknowledge that any [juror] would naturally tend to respond to what it sees as desirable — and that might easily be contrary to formal legal policy," in this case, a formal legal policy that claims a high level of certainty.[107] As an example of this phenomenon, Kaplan notes that a Southern white in the 1960s might place a lesser weight on the disutility of an erroneous conviction for an African-American and a greater weight on the disutility of erroneously acquitting an African-American defendant, leading to a lowering of the standard of proof and a higher conviction rate for African-American defendants.[108] The result, at least in this example, is that the decision theory model predicts that the applied standard of proof will vary from what the traditional view supposes.[109]

Jurors, in pursuing their own self-interests, might decide to adjust the standard of proof in other situations as well.[110] For instance, we can imagine that jurors might generally be adverse to releasing possibly guilty parties, at least with regard to certain violent crimes, such as particularly brutal murders, assaults, or rapes. If jurors were concerned that they, or others close to them, might be particularly likely to be victims of such crimes, then the jurors might overvalue the utility of a conviction. Similarly, jurors may undervalue the costs to the defendant of a conviction (accurate or erroneous), perhaps because the jurors had little experience with the personal and familial costs of imprisonment.[111] Jurors might also undervalue the disutility of a conviction because they

---

[106] *See supra* text accompanying notes 64-65.

[107] Cullison, *Probability Analysis*, *supra* note 23, at 586; *see also* Cullison, *Model of Rules*, *supra* note 23, at 240-41 (acknowledging problem that juries might not follow instructions but not resolving it).

[108] Kaplan, *Decision Theory and the Factfinding Process*, *supra* note 52, at 1075.

[109] Others have also noted the possibility that the standard of proof might vary. *See, e.g.*, Risinger, *supra* note 20, at 443-44; Posner, *Economic Approach to Legal Procedure*, *supra* note 13, at 414 (noting "the generally shared impression that the standard of proof is in fact stricter the more serious the offense").

[110] For a list of possible reasons, see Kaplan, *Cognitive Processes*, *supra* note 59, at 197, 215.

[111] *See* Tribe, *Trial by Mathematics*, *supra* note 13, at 1383-84.

fail to fully account for the marginal cost to society of punishing the defendant. So, if we assume that jurors might overvalue convictions and undervalue acquittals in some circumstances, then we are led to the conclusion that jurors — in the course of maximizing their own personal utilities — could, at least in some circumstances, end up using a standard of proof that differs significantly from the standard of proof that the traditional view suggests. More generally, decision theory predicts that the standard of proof will be high only if jurors' own values will result in a high level of certainty.[112]

Jurors may not always lower the standard of proof. Indeed, in many cases the standard of proof will still be quite high. The key is that the standard of proof may *vary* from case to case: an observation that is consistent with the empirical data. When the traditional view is viewed as a description of how jurors behave, it is attempting to model the behavior of jurors in a *particular case.* This implies that jurors should be setting the utilities based on that particular case. Thus, details about the crime, the defendant, the victim, or the penalty may affect the utility that a juror would assign to an outcome in a particular case. Sometimes the resulting burden will be quite high; other times it will be lower.

The preceding description of juror behavior is, of course, a bit simplistic, and there are at least two possible responses. First, the model assumed that jurors will seek to maximize their own personal utility, with the result that the standard of proof will not be as high as the traditional view predicts. One might challenge this assumption and argue that jurors will not maximize their own personal utility, but instead will "irrationally" assign values to those functions that will result in a high level of certainty for the standard of proof. Second, the initial description of jury behavior ignores the constraints — both formal and informal — that the legal system and society place upon jurors' behavior. One might, therefore, claim that these alternative forces act upon jurors so that they, despite their own perceived self-interest, do require a high level of certainty in all cases. In the next two sections, I discuss why these two responses actually further clarify why jurors vary the standard of proof in criminal cases.

---

[112] *See supra* text accompanying note 104.

2. Jurors As Irrational Actors

Jurors are not fully rational, and we might conclude that they may not act in a way in accord with their own personal utility.[113] For instance, Professors Daniel Kahneman and Amos Tversky, whose work lies at the heart of much of behavioral law and economics, have tried to show that expected utility theory does not accurately describe the decisions that people make.[114] In particular, they demonstrate that people make decisions not based on absolute utilities, but instead based on gains and losses from perceived reference points. Furthermore, Kahneman and Tversky also show that people evaluate decisions between two choices not based on the probabilities, but rather on "decision weights," which are usually lower than the corresponding probability.

Kahneman and Tversky's research suggests that the burden of proof that a decision maker would set would not be based on the utilities of the two choices, but rather on the "values" of the two choices.[115] In Kahneman and Tversky's theory (which they refer to as "prospect theory"), a decision maker evaluates choices (or what they refer to as "prospects") by calculating the value of each prospect ("V"), and then choosing the prospect with the higher value.[116] In the context of a typical criminal trial, there would be two possible prospect values to be evaluated: the value of a guilty verdict, V(C), and the value of a not guilty verdict, V(A).

According to Kahneman and Tversky, V can be expressed by reference to two scales, $\pi(p)$ (defined as a decision weight) and $v(x)$ (defined as the subjective value of the outcome), which are comparable, respectively, to the probability of the choice and the expected utility of that choice.[117]

---

[113] Jurors may fail to act in ways that are in accord with expected utility theory yet still are "rational." Much depends on how one defines what is "rational" and how we account for preference changes. *See, e.g.*, Richard A. Posner, *Rational Choice, Behavioral Economics & the Law*, 50 STAN. L. REV. 1551 (1998) (defining "rationality" as "choosing the best means to the chooser's ends").

[114] *See* Kahneman & Tversky, *supra* note 15, at 276-77. There are other alternatives to expected utility theory, although not all would be relevant here. *See, e.g.*, Gilboa & Schmeidler, *supra* note 69, at 607-08, 621-27 (1995) (positing theory for situations in which decision maker cannot imagine all possible outcomes and probabilities).

[115] *See* Kahneman & Tversky, *supra* note 15, at 275. In addition, bounded rationality makes it difficult to model likely behavior. For another attempt to model some of Kahneman & Tversky's data, see ARIEL RUBINSTEIN, MODELING BOUNDED RATIONALITY 25-39 (1998).

[116] In contrast, expected utility theory suggests that a decision maker should calculate the expected utility (again, "EU") for the various choices, and then decide upon the one with the highest expected utility.

[117] Kahneman & Tversky, *supra* note 15, at 275. More precisely, $\pi(p)$ is a function that

Thus, in a situation where there are two possible outcomes to a choice, y and z (both with their own utilities), Kahneman and Tversky's theory would predict that the value of that choice would be $V(x) = (\pi(p) * v(y)) + (\pi(q) * v(z))$.[118] In theory, we might then try to calculate the standard of proof by solving for the decision weight where the value of a conviction is the same as the value of an acquittal, i.e., V(C) = V(A). This leads to

$$\pi(p)\,(v(cg) - v(ag)) = \pi(1 - p)\,(v(ai) - v(ci)),^{119} \qquad (4)$$

where $v(cg)$ refers to the value of convicting the guilty, $v(ci)$ refers to the value of convicting the innocent, $v(ai)$ refers to the value of acquitting the innocent, and $v(ag)$ refers to the value of acquitting the guilty. Further solution would require us to assume that $\pi(p) + \pi(1- p) = 1$, but Kahneman and Tversky make clear that this is not true and instead posit a "subcertainty principle": $\pi(p) + \pi(1- p) < 1$.[120]

The result is that prospect theory cannot precisely describe a method by which jurors might calculate the standard of proof. Two simplifying assumptions, however, might allow for such a calculation. If we first assume that $\pi(p) + \pi(1- p) = 1$ (contrary to the subcertainty principle) and second assume that $\pi(SP) = SP$ then we get:

$$SP = \frac{1}{1 + \left(\frac{(v(cg) - v(ag))}{(v(ai) - v(ci))} \right)} \qquad (5)$$

---

assigns a decision weight to any particular probability p, and v is a function, much like the expected utility function, that assigns a "value" to a particular outcome x (in the same way that EU would assign an expected utility to a particular outcome). *Id.*

There are differences between $\pi(p)$ and p and between $v(x)$ and u that are crucial to Kahneman and Tversky's theory. For instance, the decision weight can coincide with the actual probability of an event, such that $\pi(p) = p$, but often will not do so. *Id.* at 280. This is because "[d]ecision weights measure the impact of events on the desirability of prospects, and not merely the perceived likelihood of these events." *Id.* Similarly, the value and the utility of an outcome could be the same, such that $v(x) = U(x)$, but need not do so. "An essential feature of the [prospect] theory is that the carriers of value are changes in wealth or welfare, rather than final states." *Id.* at 277. In other words, for Kahneman and Tversky, people do not assign values to things as absolutes, but rather as degrees of change from a reference point. So while the values that people assign to various outcomes could mirror the social utility that flows from an outcome, it need not do so.

[118] Kahneman & Tversky, *supra* note 15, at 276. In Kahneman and Tversky's theory, q need not, and often will not, equal $1-p$. *See id.* at 275-76.

[119] We reach this by noting that: $V(C) = (\pi(p) * v(cg)) + (\pi(1 - p) * v(ci))$ and $V(A) = (\pi(1 - p) * v(ai)) + (\pi(p) * v(ag))$. For the similar equations in expected utility theory, see *supra* note 60.

[120] Kahneman & Tversky, *supra* note 15, at 281.

which is quite similar to Equation 3, above. Thus, Kahneman and Tversky's work might suggest that jurors come to decisions about the standard of proof by relying on their *values* of the various verdict possibilities, rather than on their *utilities* for those possibilities.

Furthermore, the assumptions on which this is based may not be as radical as they initially appear. For instance, Kahneman and Tversky's own hypothetical weighting function graph suggests that it may be possible to say that $\pi(p)$ is nearly equal to p (that is, $\pi(p) \approx p$).[121] If this is true, then the second assumption is only slightly distorting. Similarly, we may be able to say that while $\pi(p) + \pi(1-p) < 1$, $\pi(p) + \pi(1-p) > .95$. If so, then perhaps $\pi(p) + \pi(1-p) = c$, where c is a constant value that almost equals 1 (that is, somewhere between .95 and 1). Going back to our first assumption, we would then get:

$$SP = c\left(\frac{1}{1+\left(\frac{(v(cg)-v(ag))}{(v(ai)-v(ci))}\right)}\right) \qquad (6)$$

This equation suggests that jurors will make decisions about the standard of proof by reference to their utilities, and then will also tend to discount (slightly) the value of the standard of proof by the constant c.

Prospect theory implies, then, that accounting for jurors' irrationality should lead to the conclusion that standard of proof actually applied by jurors is *lower*, not higher, than decision theory would suggest. Jurors will deviate from the decisions predicted by the decision theory model by substituting their own value functions for the utility functions. Jurors' decisions about the standard of proof will, therefore, depend upon their pre-existing reference points because the values can only be calculated in reference to those points. The consequence is that the jurors' standard of proof will be high in all criminal cases (as predicted by the traditional view) *only if* the jurors are assigning values that are comparable to the utilities that the traditional view relies upon. Even without empirical data, this would seem unlikely. As defined by Kahneman and Tversky, "value" is the gain or loss from a perceived reference point. If anything, it seems more likely that the "value" to a juror of an incorrect acquittal is likely to be less than its "utility," for the

[121] *See id.* at 283 fig.4.

juror would perceive a decision to acquit as "freeing" a guilty defendant. Similarly, the value of an incorrect conviction is likely to be greater than its absolute utility if the juror believes that the innocent defendant still poses a danger to the juror or to her family or friends. Relying on "values" rather than on "utilities" should lead to a lower, not a higher, standard of proof.[122]

Furthermore, the inclusion of the constant *c* might also lower the standard of proof applied by jurors. Assuming that jurors were to actually use *c* in their calculation of the standard of proof, we need a value for it, which should be less than, but quite close to, 1. The result is that jurors (if they are making the simplifying assumptions I have suggested) would calculate the standard of proof at a number that is slightly lower than what the decision theory model would suggest, even if the jurors' "values" were the same as the utilities suggested by the traditional view. Prospect theory, therefore, suggests a lower, not a higher, standard of proof than we generally assume.

3. Constraints on Jurors

The second response to the expected utility model of juror behavior is that jurors require a high level of certainty because the legal system requires jurors to do so, through formal and informal constraints on juror behavior. For example, a formal constraint on juror behavior would be jury instructions, which tell the jurors what standard of proof they are to apply in a case. Of course, this constraint is effective only if jurors obey the instruction. So alternatively, jurors might be constrained in their behavior by other, non-legal forces. This presupposes that such forces have an effect on jurors that is powerful enough to overcome jurors' tendency to maximize their own self-interest. The actual operation of the criminal justice system, though, suggests neither formal nor informal constraints have a large effect on jurors' decisions about the standard of proof.

The formal constraints on jury behavior are surprisingly small: jury instructions are really the only ones. Other formal constraints are imaginable. Courts could review the decisions by jurors about the standard of proof and give them incentives to make correct decisions by rewarding them for correct decisions and/or punishing them for incorrect decisions. The Anglo-American legal system, however, has

---

[122] *See also* Tribe, *Trial by Mathematics*, *supra* note 13, at 1383-84 (claiming, based on cognitive dissonance, that jurors will do something similar).

generally been resistant to the idea of punishing jurors for "incorrect" verdicts, and it has never been seriously suggested that we should reward jurors for "correct" verdicts.[123] Another approach might be to review the decisions of jurors on the standard of proof and reverse their decisions where we believe that they have gotten the decision wrong. The prohibition on double jeopardy, though, forbids any review of acquittals, which would mean that we could only review convictions where the jurors applied the wrong standard of proof — presumably a standard of proof that was too low.[124] Even this limited approach is forbidden in our legal system, though, because it would require impeaching the jury's verdict.[125]

Jury instructions in general, and the reasonable doubt jury instruction in particular, are also fairly ineffectual at constraining juror behavior. These instructions, on their face, state that the jury must follow the law as it is given to them and also tell them that they must apply the reasonable doubt standard. Our formal ideology assumes that jurors, having been so instructed, then apply a high level of certainty. But the realities of the courtroom suggest that this is not so. In most places, the jurors still receive jury instructions only orally and are not given written instructions, even if they ask for them.[126] In many jurisdictions, therefore, a juror has only one moment during which she is even exposed to instruction: an oral presentation by the judge.[127] This exposure occurs at the end of a case that has frequently lasted for several days and in the midst of an oral statement from the judge on numerous different issues.[128] So there is good reason to suspect that jurors, under

---

[123] THOMAS ANDREW GREEN, VERDICT ACCORDING TO CONSCIENCE: PERSPECTIVES ON THE ENGLISH CRIMINAL TRIAL 1200-1800, at 200-64 (1985) (discussing limited history of fining and imprisoning jurors). For a description of the events leading up to *Bushell's Case*, which established this rule, see HANS & VIDMAR, JUDGING THE JURY, *supra* note 101, at 21-23.

[124] *See, e.g.*, Fong Foo v. United States, 369 U.S. 141 (1962) (holding that even acquittal based on egregiously erroneous foundation cannot be set aside).

[125] The decision about the standard of proof is necessarily a subset of the decision to convict or acquit. Existing law makes it impermissible to inquire into how the jury reached its verdict. *See* Nancy J. King, *Juror Delinquency in Criminal Trials, 1976-1996*, 94 MICH. L. REV. 2673, 2720-21 (1996).

[126] *See* JEFFREY ABRAMSON, WE, THE JURY 91 (1994); Laurence J. Severance & Elizabeth F. Loftus, *Improving the Ability of Jurors to Comprehend and Apply Criminal Jury Instructions*, 17 LAW & SOC'Y REV. 153, 155 & n.4 (1982).

[127] Merely giving the instruction twice, once at the beginning of the case and once at the end of the case, might improve comprehension. *See* HANS & VIDMAR, JUDGING THE JURY, *supra* note 101, at 122-23; Elwork et al., *supra* note 103, at 36.

[128] The median federal criminal trial in fiscal year 1997 took two days. *See* BUREAU OF JUSTICE STATISTICS 1997, *supra* note 35, at 417. The mean time, however, was at least 2.74,

these conditions, cannot and do not internalize the content of the instructions.[129]

Furthermore, even if jurors were to receive the oral instructions in more idealized conditions, the instructions are unlikely to be understood because they are so poorly drafted.[130] This is certainly true of the reasonable doubt instruction.[131] Recall the instruction quoted in the Introduction and consider this example, from another prominent federal pattern jury instruction treatise:

> A reasonable doubt is a doubt based upon reason and common sense — the kind of doubt that would make a reasonable person hesitate to act. Proof beyond a reasonable doubt must, therefore, be proof of such a convincing character that a reasonable person would not hesitate to rely and act upon it in the most important of his or her own affairs.[132]

These instructions appear to do little to educate the jurors in what is meant by proof beyond a reasonable doubt. They certainly do not quantify how much proof is needed to convict a defendant. But what is most remarkable is how vague and opaque they are. These instructions used in the federal system talk about hesitating to act in an important matter. What does that mean? State instructions are no better.[133] They

---

and probably higher. *Id.* Furthermore, over 27% of all federal criminal trials lasted at least four days. *Id.*

[129] *See generally* ABRAMSON, *supra* note 126, at 91 (arguing that jurors are unlikely to understand difference between murder and manslaughter based on judge's instructions).

[130] *See id.* ("To anyone who has ever witnessed a judge instructing a jury, it is clear that our system does not even pretend that the instructions are meaningful."); HANS & VIDMAR, JUDGING THE JURY, *supra* note 101, at 121; Darryl K. Brown, *Regulating Decision Effects of Legally Sufficient Jury Instructions*, 73 S. CAL. L. REV. 1105, 1113 (2000) ("Use of instructions that are legally sufficient but greatly deficient in clarity and positive decision effect is widespread and well documented."); Elwork et al., *supra* note 103, at 35-36; Lind, *supra* note 6, at 27-29; Severance & Loftus, *supra* note 126, at 157-61; Walter W. Steele & Elizabeth G. Thornburg, *Jury Instructions: A Persistent Failure to Communicate*, 67 N.C. L. REV. 77, 88-95 (1988) (demonstrating that several different common jury instructions could be rewritten to create greater juror understanding).

[131] *See, e.g.*, Bradley Saxton, *How Well Do Jurors Understand Jury Instructions? A Field Test Using Real Juries and Real Trials in Wyoming*, 33 LAND & WATER L. REV. 59, 97 (1998) (finding that significant numbers of jurors could not even accurately state which party, government or defendant, bore burden of persuasion); Severance & Loftus, *supra* note 126, at 174-75, 179-83.

[132] 1A O'MALLEY ET AL, *supra* note 29, § 12.10, at 168.

[133] Consider these two examples, from California and New York, respectively:

> Reasonable doubt is defined as follows: It is not a mere possible doubt; because everything relating to human affairs is open to some possible or imaginary doubt. It is that state of the case which, after the entire comparison and

talk about proof beyond a reasonable doubt as "an abiding certainty" or what occurs after a "careful and honest review" of the evidence. What is a juror supposed to draw from these statements? And finally, some jurisdictions decline to define reasonable doubt at all, leaving it to the jurors to define it for themselves.[134] It is, therefore, no surprise that these instructions have come under judicial and academic attack for failing to set forth the standard of proof in either an accurate or understandable fashion.[135] Furthermore, as I have already noted, there is no way to enforce the instructions' command as to the proper standard of proof.[136]

> consideration of all the evidence, leaves the minds of the jurors in that condition that they cannot say they feel an abiding conviction of the truth of the charge.

1 CALIFORNIA JURY INSTRUCTIONS: CRIMINAL Instr. 2.90, at 117 (1996).

> Reasonable doubt, however, is no mere whim, guess or surmise, or feeling, perhaps, that an accused person may not be guilty; nor can you avail yourself of it as a subterfuge to which to resort in avoiding a disagreeable duty. For a doubt to be reasonable, it must be one that a reasonable man or woman would have after a careful and honest review and consideration of all of the evidence in the case. It must be found in reason.
>
> This rule of law does not mean that the prosecution must show that the defendant is guilty beyond all doubt. But, at the same time, just a bit or a wisp of proof, just some testimony entered into the record which would seem to indicate the commission of the offense charged, will not do the job. Before the district attorney can produce a case justifying your finding a verdict of guilty, his (her) evidence must be firm, convincing, believable and complete enough to leave no reasonable doubt of the defendant's guilt in your mind.

1 LEVENTHAL, *supra* note 29, § 4:66, at 213.

[134] *See* SAND ET AL., *supra* note 5, at 4-9 to 4-10 (noting that Fourth and Seventh circuits decline to define reasonable doubt); Solan, *supra* note 9, at 115-16 (also noting that Fourth and Seventh circuits decline to define reasonable doubt).

[135] *See, e.g.*, Newman, *supra* note 9, at 982-84 (noting ambiguity of "hesitate to act" language in reasonable doubt instruction); Power, *supra* note 9, at 97-102 (reviewing literature critiquing juror comprehension of existing instructions); Solan, *supra* note 9, at 112-16 (arguing that most existing instructions fail to instruct jurors properly in burden of persuasion); W.C. Thompson et al., *Death Penalty Attitudes and Conviction Proneness: The Translation of Attitudes into Verdicts*, 8 LAW & HUM. BEHAV. 95, 98 (1984) ("[T]he standard of reasonable doubt is only vaguely defined. Hence jurors largely determine for themselves their threshold of conviction.") (footnote and citations omitted); *Reasonable Doubt*, *supra* note 10, at 1964-67 (arguing that existing reasonable doubt instructions serve only to confuse jurors); *see also* Cullison, *Model of Rules*, *supra* note 23, at 237 (noting this phenomenon and seeming to support it). *But see* Cohen, *supra* note 10, at 697-98 (claiming that present pattern instructions simply and clearly define reasonable doubt). Finally, some have suggested that the courts should adopt the approach of those that have refused to give any definition of reasonable doubt. *See* United States v. Hall, 854 F.2d 1036, 1044 (7th Cir. 1986) (Posner, J. concurring); *Reasonable Doubt*, *supra* note 10, at 1955.

[136] *See supra* note 123 and accompanying text.

As a result, it seems quite a stretch to imagine that our present jury instruction practice leads to the use of a high level of certainty by jurors.[137]

Even if jury instructions do not require a high level of certainty, perhaps there are other, less formal, constraints that do. For instance, attorneys may constrain jurors' selection of a decision criterion when they address the jury during closing arguments. Defense attorneys, who are seeking the acquittal of their client, certainly would seem to have an incentive to stress to the jury that the reasonable doubt standard requires a high level of certainty.[138] Prosecutors, though, have every incentive to minimize the level of certainty required, and there is good reason to suspect that they do emphasize that not all doubts need to be removed.[139] But jurors no doubt understand prosecutors and defense attorneys to be merely agents of adversarial parties, and the jurors presumably discount whatever information they receive from the attorneys in general as biased. Indeed, if jurors were likely to worry about attorneys taking actions against them, they are going to worry only about one of the attorneys: the prosecutor. Of course, the judges' instructions on reasonable doubt might help to reinforce the arguments of defense counsel that the reasonable doubt standard requires that the jurors have a great deal of certainty before convicting, and in at least some (and perhaps many) cases, jurors will generally set the burden of persuasion

---

[137] Of course, some jurors might mistakenly believe that judges, because they are agents of the criminal justice system, do have the power to sanction them. These people are probably few in number. Judges' instructions may help reinforce the arguments of attorneys, though, a subject I turn to in the next paragraph.

[138] *See, e.g.*, F. LEE BAILEY & HENRY B. ROTHBLATT, SUCCESSFUL TECHNIQUES FOR CRIMINAL TRIAL 257-58 (1971) ("Be sure that each member of the jury knows the vital difference between the rule of preponderance of the evidence in a civil case and reasonable doubt in a criminal case."); *id.* at 304 ("Your entire summation is based upon one premise: the failure of the prosecutor to establish the guilt of the defendant beyond a reasonable doubt."); THOMAS A. MAUET, FUNDAMENTALS OF TRIAL TECHNIQUE 312-314 (3d ed. 1992) (setting forth mock defense argument that jury should acquit if it has "a doubt"); JACOB A. STEIN, CLOSING ARGUMENT: THE ART & THE LAW § 208 (1996) ("The difference between the civil burden of proof of a preponderance of the evidence and the much more difficult reasonable doubt standard used in criminal cases, may be discussed. Define what is a reasonable doubt. You may say that to convict the defendant, the jury must be morally certain that the defendant is guilty beyond a reasonable doubt. Also, counsel may want to emphasize the terrible injustice of convicting an innocent person."); *see also* Dane, *supra* note 83, at 142.

[139] *See, e.g.*, SAM SCHRAGER, THE TRIAL LAWYER'S ART 81 (1999) (giving example of argument where jury was told "[d]on't sit back there and create doubts"); *id.* at 121 (noting that jurors do not leave their common sense outside courtroom); *see also* Dane, *supra* note 83, at 142.

high, or at least higher than they would if they were not instructed.[140] Nonetheless, given the conflicting information that jurors receive about reasonable doubt from prosecutors and defense attorneys and the absence of any reason for jurors to feel bound by this information, we should suspect that these arguments have much less effect than we commonly assume.

Although formal and informal mechanisms may fail to constrain jurors' selection of a decision criterion, perhaps society, in the form of social norms independent of the legal system, requires a high level of certainty. Judge Richard Posner, for one, has suggested this possibility.[141] Such social norms do affect juror behavior on other issues, such as why people bother to appear for jury service at all.[142] But even in this situation, the social norm is only partially successful. While some people show up and serve as jurors, many people do not do so — they either do not show or they lie their way out of service.[143] Similarly, the legal system does little, as a matter of formal law, that actually requires jurors to attempt to reach an accurate verdict — that is, the legal system gives jurors little incentive to expend much effort on achieving the right

---

[140] For a further discussion of this point, *see infra* Part V.

[141] United States v. Hall, 854 F.2d 1036, 1044 (7th Cir. 1988) (Posner, J., concurring) (arguing that because beyond a reasonable doubt is "deeply entrenched in the popular culture," attempts at defining reasonable doubt as probabilistic measure are apt to mislead jurors).

[142] Service on a jury usually entails opportunity costs for jurors. Because the actual pay for jury service is typically quite low (state rates vary from $5 per day to $50 per day, see *Jury Pay Scales*, DALLAS MORNING NEWS, Oct. 24, 2000, at 8A), a purely self-interested juror would presumably prefer to be doing something, almost anything, else. (I am making the assumption that jury service is not generally preferred by jurors for noneconomic reasons.) The formal legal response is to make jury service compulsory, but the punishments for failing to respond to a summons are typically quite low and the likelihood of punishment for those who do not show up is even lower. *See, e.g.*, 28 U.S.C. § 1864(b) (setting maximum sanction in the federal system at $100 fine and/or 3 days imprisonment); King, *supra* note 125, at 2700-05 (describing rare use of enforcement powers and preference for attempts to make juror service more appealing).

One potential explanation for why at least some people show up can be found in social norm theory. In recent years, commentators have focused on how society influences the behavior of individuals not just by formal legal rules but also by informal rules enforced by nonlegal sanctions. *See, e.g.*, Richard H. McAdams, *The Origin, Development, and Regulation of Norms*, 96 MICH. L. REV. 338, 350 (1997); ERIC A. POSNER, LAW AND SOCIAL NORMS (2000). Since the obligation to show up and to serve on a jury is not primarily enforced by legal sanctions, it must be enforced by a nonlegal norm. *See* Ann E. Carlson, *Recycling Norms*, 89 CAL. L. REV. 1231, 1235 (2001). *But see* McAdams, *supra*, at 348 n.43 (suggesting that individuals show up for jury service because of legal obligation).

[143] In some major metropolitan areas, the response rate can be as low as 10-20%. *See* Mark Curriden, *No Excuses: New Yorkers Who Try to Avoid Jury Duty Find that System Has Gotten Serious About Service*, DALLAS MORNING NEWS, Oct. 24, 2000, at 1A.

outcome.[144] Again, the reason that jurors do expend a great deal of effort in resolving cases correctly is because there is a societal expectation that they do so. But here too, jurors will on occasion deliberate only briefly, and will generally deliberate only until they have sufficient votes to return a verdict.[145] So while a social norm of a high level of certainty may constrain juror behavior in some circumstances, we should also expect that in other circumstances the social norm will be overridden by the jurors' own self-interests.

The result is that we should be skeptical that a social norm of "require a great deal of certainty before convicting" will be fully effective in constraining juror behavior. That is not to say it will have no effect. Just as some people do show up for jury service and many jurors do try to reach the right outcomes, some jurors will be constrained by the social norm. What we should not expect, however, is that jurors will always set aside their own self-interests and always require a high level of certainty in all cases. To the contrary, jurors seem likely, at least in some cases, to vary the burden and to require substantially less certainty. This is true even if we assume that jurors are constrained by various formal and informal norms. Thus, it should come as no surprise that jurors do, in fact, vary the standard of proof.

III. THE PROBLEM OF UTILITIES: THE FAILURE OF THE TRADITIONAL VIEW AS A NORMATIVE MODEL

The failure of the traditional view as a model of how jurors behave is not by itself fatal to its usefulness. Kaplan and Cullison themselves were never clear about whether they were describing how the standard of proof is set or instead how the standard of proof *ought to be* set. If the latter, normative view of their model is accepted, then what the traditional understanding tells is that society should want a high level of certainty in all criminal cases because such a result will result in greater

---

[144] The closest we come is to require that jurors take an oath to try the case truly.

[145] *See* Annotation, *Effect on Verdict in Criminal Case of Haste or Shortness of Time in Which Jury Reached It*, 91 A.L.R.2d 1238, 1239-40 (1963) (reporting that "no criminal case has been found in which haste or shortness of time taken by a jury in arriving at its verdict was held to amount to reversible error"). Indeed, deliberation times as little as three, six, seven and eight minutes have been affirmed. *See* Later Case Service, 91 A.L.R.2d 1250 at 24. The empirical evidence suggests that jurors generally stop deliberating as soon as they obtain sufficient numbers for a verdict. *See, e.g.*, Norbert L. Kerr et al., *Guilt Beyond a Reasonable Doubt: Effects of Concept Definition and Assigned Decision Rule on the Judgments of Mock Jurors*, 34 J. OF PERSONALITY & SOC. PSYCHOL. 282, 292 (1976). In other words, jurors sacrifice further deliberation (which might be supposed to create greater accuracy) in favor of a quicker resolution.

social utility. On this reading of the traditional view (as modified by Kaplan and Cullison's model), the empirical evidence does not undermine the theory; it instead suggests the need for reform.

Unfortunately, even when the traditional understanding is viewed as a normative theory for how the standard of proof ought to be set, it turns out to be unsatisfying. The decision theory model fails to provide any method by which to determine the various utilities it considers. Furthermore, existing theories of criminal law do not provide the sort of assistance that could lead to any consistent predictions about the appropriate standard of proof.

A. Decision Theory and Utilities

For Kaplan and Cullison's model to function as a normative theory of reasonable doubt, we need to know the relative values of verdicts. In the traditional view, this was solved by assuming a large difference between the costs of erroneous convictions and erroneous acquittals: something along the line of 10:1 (or 5:1 or 20:1). As Equation 2 demonstrates, using these values results in a high standard of proof (slightly in excess of .90). So long as there is a rough societal consensus that this is the correct ratio, there is no problem reaching the conclusion that a high level of certainty is appropriate for all cases.

The difficulty comes when, as discussed in Part I.C, accurate verdicts have to be taken into account. At this point, relative values need to be assigned for two new variable functions: the utility of an accurate conviction and the utility of an accurate acquittal. And here, no rough societal consensus appears to exist about how to value the utilities, or at least no consensus that would apply to *all* cases (I will argue in Part IV that there should be a consensus that arises in individual cases). Indeed, as noted above, there are a whole host of different ways to order these utilities, all of which lead to different conclusions about the standard of proof: very high or relatively low or somewhere in between.[146]

So is there a way to assign the necessary utilities? Decision theory cannot supply them. The expected utility model provides us a heuristic about how parties ought to choose among alternatives under a condition of risk.[147] For instance, we can evaluate whether it is better to take a 10% chance for $1000 or a 50% chance for $100. All other things being equal, the expected utility model suggests that the first choice here is best

---

[146] *See supra* text accompanying notes 61-68.

[147] *See* Paul J.H. Schoemaker, *The Expected Utility Model: Its Variants, Purposes, Evidence & Limitations*, 20 J. ECON. LITERATURE 529, 530 (1982).

because the expected utility of that choice is $100, whereas the expected utility of the second choice is only $50. This example, though, depends for its success on our implicit acceptance of the relative values of $1000 and $100: that $1000 is ten times more valuable to us than $100. But what if my preferences are not so linear; what if, instead, I really need $100 right now to pay a bill, and that an extra $900 would be nice, but not ten times as important as getting that $100 right now, such that $1000 is really worth only twice as much to me as $100?[148] In such a case, expected utility theory would suggest that I should prefer the latter choice, because the expected utility *to me* of that choice is higher.

This insight is crucial to behavioral economists' critique of standard expected utility models of human behavior, but its importance goes beyond that narrow realm. Decision theory models themselves do not (and cannot) provide the utilities that are central to their analyses. Instead, these values must come from external sources. For many choices, such as the example I started with above, the values are generally agreed upon. But for many other choices, such as the value of verdicts, there may not be any clear consensus.

In deciding how to set the standard of proof for criminal cases, one approach might be to start from thinking about the purposes of the criminal law and, in particular, the criminal sanction. Once we know why we are imposing criminal sanctions on guilty defendants, we could perhaps then evaluate the utilities and disutilities of the various verdict possibilities. This, in turn, would lead to the setting of the standard of proof for all criminal cases. The difficulty is that there is no clear consensus for our use of the criminal sanction. Theories include deterrence, incapacitation, expression, retribution, and rehabilitation.[149] In the next section, I investigate what assistance, if any, deterrence theory can give us to setting the relevant utilities, before exploring other theories in the remaining sections.

B. Deterrence and Utilities

One seemingly logical candidate for assigning values to utilities in a decision theory model might be Gary Becker's deterrence theory, which starts with the claim that crimes are committed because the expected utilities of the activity outweigh the expected disutilities to the

---

[148] *See* Kahneman & Tversky, *supra* note 15.

[149] *See* JOHN KAPLAN ET AL., CRIMINAL LAW: CASES & MATERIALS 33-84 (4th ed. 2000); WAYNE R. LAFAVE & AUSTIN W. SCOTT, JR., CRIMINAL LAW § 1.5(a) (2d ed. 1986).

individual.[150] If apprehending criminals and imposing punishments were costless, then presumably society would maximize its own welfare by making it as certain as possible that criminals would be apprehended and punished, because this would result in the greatest disincentive for individuals to engage in crime. In reality, however, the costs of crime are not just a function of the loss from crimes themselves, but also the costs of apprehension (and trial) of criminals, the likelihood of punishment, and the costs of punishment to society (and perhaps to the criminal).[151] The net loss to society is, therefore, a function of three terms: the harm caused by the crime, the cost of apprehending and punishing criminals, and finally the total social loss from punishments.[152]

Becker assumes that the only two variables are the probability of detection and conviction (p) and the amount of punishment (f).[153] We can assume for purposes of this Article, however, that the only variable is p, because the amount of the punishment for the crime has already been set by the substantive criminal law.[154] Finally, there is some optimal probability of apprehension and punishment, designated p*, whereby social loss is minimized.

The question, then, is what are the utilities and disutilities associated with the various verdict possibilities, given the stated desire to optimize the probability of apprehension and punishment (again, p)? Consider, for instance, accurate convictions, which might seem a simple calculation

---

[150] Gary S. Becker, *Crime and Punishment: An Economic Approach*, J. POL. ECON. 169, 176-77 (1968); *see also* POSNER, ECONOMIC ANALYSIS, *supra* note 13, at 242. This expected utility calculation is dependent upon the likelihood that the defendant will be punished, such that EU(CR) = pU(Y – f) + (1 – p) U(Y), where EU(CR) is the expected utility of crime, p is the probability of punishment, U(Y) is the net utility to the individual of the criminal action and U(Y – f) is the net utility to the individual of criminal action and punishment, because f is the amount of punishment for the crime. Becker, *supra*, at 176 n.16; DAVID J. PYLE, THE ECONOMICS OF CRIME AND LAW ENFORCEMENT 93 (1983); *see also* Samuel Cameron, *The Economics of Crime Deterrence: A Survey of Theory and Evidence*, 41 KYKLOS 301, 302 (1988).

[151] *See* Becker, *supra* note 150, at 180-85. This can be represented as L = D(O) + C(p, f) + bfpO, where L is the net loss to society from crime, D is the net harm to society from the crime, O is the number of crimes committed, C is cost of apprehension and conviction of criminals, p is the probability that a criminal will be punished, f is the amount of punishment for the crime, and b is the ratio of the costs of punishment to society and the individual. *Id.* at 181.

[152] The total social loss from punishments is represented in the equation by the term bfpO.

[153] *See* PYLE, *supra* note 150, at 90 (describing Becker's model).

[154] That is, the amount of punishment for a particular crime is normally set by statute. Of course, most criminal statutes set ranges of punishment, not exact punishments. Nonetheless, for my purposes here we may assume that f tends to be relatively invariable, particularly in "guidelines" systems, such as the federal one, where the amount of punishment is set by an elaborate calculation.

under deterrence theory. The utilities of accurate conviction will depend upon how such convictions affect p. If the effect of an accurate conviction is to move p toward p*, then U_{cg} is positive. But if the effect of an accurate conviction is to move away from p*, then U_{cg} is negative, because it is decreasing social welfare (the cost of obtaining the additional conviction outweighs the benefit). So, to know U_{cg}, we have to know whether a criminal statute is being over-enforced or under-enforced in order to decide whether enforcement in a particular case will increase or decrease social welfare.[155]

In addition, the utility of an accurate conviction will also depend on the rates of apprehension of criminals and the rates at which they plead guilty.[156] This is because p is really dependent not just on the probability of an apprehended defendant eventually being convicted at trial, but also the rates at which criminals are apprehended and the rates at which they plead guilty. So, to determine whether an accurate conviction is moving p toward or away from p* (assuming we know p*), we would need to know both the rates at which criminals are apprehended and plead guilty. Without knowing those rates (and we know very little about them), we cannot know whether a particular accurate decision to convict would increase or decrease social welfare.[157]

---

[155] Connolly makes a related point in noting that, in general, one cannot calculate *SP* (which he refers to as *P**) without information about the ratio of guilty and innocent defendants who are tried. *See* Connolly, *supra* note 47, at 104-05.

[156] The value of p is actually dependent on several subvariables, such that $p = p_a . (p_{gp} + p_c)$, where p_a is the probability of apprehension, p_{gp} is the probability of an apprehended defendant entering a guilty plea, and p_c is the probability of an apprehended defendant eventually being convicted at trial. Note that p_c is set in relation to all cases, such that even if the probability of being convicted at trial is .8, if the total percentage of defendants who go to trial is 10%, then $p_c = .08$. Say that only 15% of all cases go to trial. Then, even if the likelihood of conviction at trial was 100%, the maximum value of p_c would be .15. The result is that deterrence will come mainly from apprehension and guilty pleas, not trial results. ZUCKERMAN, *supra* note 11, at 128.

[157] Here, I am assuming that the cost of the sanction to the offender is a direct cost to society. *See, e.g.*, Louis Kaplow & Steven Shavell, *Fairness Versus Welfare*, 114 HARV. L. REV. 961, 1226 (2001) (asserting that social welfare depends on costs of punishment). *But see* Kenneth G. Dau-Schmidt, *An Economic Analysis of the Criminal Law As a Preference-Shaping Policy*, 1990 DUKE L.J. 1, 11 (1990) (arguing that including criminal benefits in concept of social welfare defies common sense). One objection to this might be that there would be no alteration in the costs to society. Say that the cost imposed on the defendant is imprisonment. To obtain the same deterrence in my reasonable doubt regime, I have suggested that the cost of the sentence to the offender must be five times more than it would be in my preponderance of the evidence regime. But because only 1/5 as many defendants would be punished, the total cost would be the same. This would be true, were it not for the discount rate on sentences. It turns out that the most likely result, though, is that the cost of imprisonment to defendants rises at a lower rate than the term itself. *See* A. Mitchell Polinsky & Steven Shavell, *On the Disutility and Discounting of Imprisonment and the*

So far the model has assumed (inaccurately, I believe) that the other components of the cost of crime, such as the social costs of crime, the costs of apprehension (and trial) of criminals, and the likelihood of punishment, are fixed. The value of each of these, however, should vary depending on the particular crime we are discussing and the punishment attached to it. That is, for particular crimes, the social costs will be higher or lower, the costs of apprehending and punishing defendants will be higher or lower, and the punishments will impose differing costs, both on the defendant and on society. Accordingly, while we could generalize a value for p^*, the reality is that p^* should vary depending on the particular crime. This implies that the utility of an accurate conviction depends on the type of crime and punishment involved in a particular case.

Thus far, I have only discussed the difficulty of coming to a decision about the utility of an accurate conviction. Coming to some normative conclusion about the utility and disutility for the other three possibilities (accurate acquittals, inaccurate convictions, and inaccurate acquittals) is even more difficult. Consider the (dis)utility of erroneous convictions. Erroneous convictions impose a number of costs. First, they impose the same costs that all convictions impose: the costs on society of punishing the defendant but without any corresponding benefit. Second, an erroneous conviction creates costs because it decreases the deterrent effect of the punishment. This is because it has the effect of *decreasing* the effect of a punishment. If an individual already knows that she will run some chance of being punished regardless of whether or not she engages in the activity, the cost of the sanction decreases.[158] Therefore, erroneous convictions impose costs not only because we have wasted the costs of punishment on an individual who should not be punished, but we have

---

Theory of Deterrence, 28 J. LEGAL. STUD. 1, 5, 9-10 (1999). Therefore, the actual amount of imprisonment time would have to rise not to five times that of the preponderance regime, but perhaps to six or seven or more times. The overall result is more costs for society. But if defendants do not discount the cost of imprisonment, then it is possible that raising the standard of proof would create a social benefit. *See id.* at 8.

[158] The cost of punishment to an individual is $pU(Y - f)$. *See supra* note 150. Now, think back to Table 1. There, individuals had the exact same probability of punishment (because the amount of evidence generated against both types of defendants was the same). Theoretically, a conviction against a defendant would generate the following expected utility: $EU(CR) = pU(Y - f) + (1 - p)\ U(Y) + qU(f)$, where q is the probability of being punished when innocent. Because $p = q$, and assuming that $U(Y - f) = U(Y) - U(f)$, then the expected utility of engaging in the criminal behavior is simply $U(Y)$. The defendant is aware that she runs a 10% chance of being convicted whether or not she engages in the criminal activity, so the probability of conviction has no effect on her. *See* POSNER, ECONOMIC ANALYSIS, *supra* note 13, at 605 n.1.

also undermined the deterrent effect of punishment. In the absence of any clear information about the relative utilities (or disutilities) of accurate and erroneous convictions, it is hard to imagine that deterrence theory will provide much insight.[159]

C. Social Norms and Utilities

This is not a problem specific to deterrence theory but rather can be generalized to any theory of criminal law. Consider, for instance, expressive theories of criminal law, which have been advocated recently by law and social norms scholars. One such theory is that of Dan Kahan, who has recently described how the criminal law may be viewed for its expressive function.[160] According to Kahan, the criminal law, and more particularly, criminal punishment, serves to signal society's disapproval of an individual's actions. Behavior is criminalized, in part, in an effort to express society's moral condemnation of the behavior, as well as the values that the behavior symbolizes.[161] In addition, behavior comes to be criminalized as part of a competition between competing groups *within* a society. In a society where groups have competing moral values, those groups will inevitably attempt to use the criminal law to demonstrate

---

[159] Tribe notes that information about "how much deterrent effect will flow from" convictions (among other pieces of information) should be considered in determining a standard of proof in criminal cases. *See* Tribe, *Trial by Mathematics*, *supra* note 13, at 1385 n.174. But Tribe erroneously believes that such information could not "affect the decision in the Kaplan-Cullison model." *Id.* To the contrary, the model could incorporate such information through the utility functions. The problem is that we do not have the information to set those values.

[160] Dan M. Kahan, *The Secret Ambition of Deterrence*, 113 HARV. L. REV. 413 (1999) [hereinafter Kahan, *Secret Ambition*]; *see also* Dan M. Kahan, *Social Meaning and the Economic Analysis of Crime*, 27 J. LEGAL STUD. 609 (1998) [hereinafter Kahan, *Social Meaning*]; Richard H. McAdams, *An Attitudinal Theory of Expressive Law*, 79 OR. L. REV. 339 (2000) [hereinafter McAdams, *An Attitudinal Theory*]. Scott Sundby may also fall into this group, at least to the extent of agreeing that the criminal law is the result of the community's expression of what constitutes criminal behavior. *See* Sundby, *supra* note 12, at 490, 494 (stating that "[a] verdict of guilty . . . is not simply an administrative decision to lock the defendant up, but a factual finding that an individual engaged in behavior that society — through the legislature — has decided is neither justifiable nor excusable" and that "[w]hen a jury is convened it becomes the means of carrying out the particular community's judgment of what currently merits criminal sanctions"). For a more general description of one approach to expressive theories of law, see Elizabeth S. Anderson & Richard H. Pildes, *Expressive Theories of Law: A General Restatement*, 148 U. PA. L. REV. 1503 (2000). For an argument that none of these approaches is a true "expressive" theory of law, see Matthew D. Adler, *Expressive Theories of Law: A Skeptical Overview*, 148 U. PA. L. REV. 1363, 1495 (2000).

[161] Kahan, *Secret Ambition*, *supra* note 160, at 420. As an example, theft is criminalized, but not economic competition, because society wishes to express its condemnation of theft but not economic competition. *Id.*

that their values, and not those of other groups, are predominant.[162] Thus, criminal law also serves as a way for one subset of society to express its triumph over other groups.[163]

These expressive theories have some initial attraction because they tell us more about the utilities of convicting and acquitting defendants independent of the probability of convictions and acquittals. By extending the expressive theories, we might show that accurate convictions and acquittals generate utility because they properly express society's disapproval of certain behaviors and approval of other behaviors. Erroneous convictions and acquittals have negative utilities because they are sending the wrong message. Inaccurate verdicts send a message that society either approves of behavior it does not really approve of or a message that it disapproves of behavior that it actually approves of, at least if society becomes aware of the inaccurate nature of the verdict.[164] Therefore, it might be possible to measure the expressive effect of each type of verdict and come to a conclusion about the proper standard of proof.

There are several problems with such an approach. First, this extension of expressive theories assumes that the public at large is aware of the decisions of criminal juries. But the public is generally not aware of most murder trials in the United States, let alone any significant

---

[162] *Id.* at 422.

[163] Other theorists have taken similar tacks. Richard McAdams, for example, has claimed that the law (and presumably the criminal law) can affect change in people's behavior by altering people's beliefs about which behaviors will generate approval from society. McAdams, *An Attitudinal Theory*, *supra* note 160, at 340. McAdams begins by assuming that people care about how others will react to their behavior and that, in general, people wish to have the approval of others for the actions that they have taken. The actions of governmental entities can then alter people's behavior by changing their beliefs about what actions are approved of (or disapproved of) by other members of society. Thus, a legislature's actions to outlaw a behavior may signal to people that the members of that entity believe that society as a whole disapproves of the behavior. People then will change their behavior such that they will be less likely to engage in the prohibited activity (in accordance with the legislation) not just because they are worried about the sanction from violating the law, but because people have altered their assessments of how likely it is that other people will disapprove of the activity.

[164] To the extent that there is no awareness of the inaccurate nature of an erroneous verdict, there might actually be positive expressive utility associated with an erroneous verdict. The public, unaware of the mistake, assumes that guilty people are being punished and innocent persons are freed. Accounting for this possibility (while at the same time offsetting the negative utility if the verdict is later discovered to be inaccurate) only makes it more difficult, not less, to come to any conclusion about the proper normative value for the standard of proof because it would require calculating the possibility that an erroneous verdict will be discovered.

portion of the verdicts in other criminal cases.[165] Thus, in the vast majority of criminal cases (with a few exceptions that I will discuss below), there is little expressive utility, either positive or negative, from a jury verdict (guilty or not guilty, accurate or erroneous).[166]

Second, and more importantly, expressive theories do not appear to give any independent normative significance to a particular verdict. Instead, a verdict will be viewed as accurate or inaccurate simply to the extent to which it mirrors the purposes of society in permitting or forbidding certain actions. Decisions about which actions are approved — or are not approved — by society are judged simply by whether society does in fact approve or not approve of those actions; no independent guide to which activities society should or should not approve exists.[167] Instead, the best that an expressive theory may do is to give us an insight into what values society places on the various utility functions.[168] These values, in turn, are likely to vary considerably from

---

[165] The number of criminal cases of which the public is aware is truly miniscule. In 1998, there were nearly 6,000 defendants tried in the federal system alone. BUREAU OF JUSTICE STATISTICS 1998, *supra* note 35, at 410-11 (3,629 convictions and 1,081 acquittals). In 1997, 1,322,201 persons were charged with serious offenses in state courts, with an overall total of 7,634,815 arrested. *Id.* at 341 table 4.9. In 1996, the number of felony cases filed was close to a million. *See* NORMAN ABRAMS & SARA SUN BEALE, FEDERAL CRIMINAL LAW AND ITS ENFORCEMENT 13 (3d ed. 2000). Assuming approximately 1 million serious criminal cases, and if only 2% of the serious cases go to trial, that means 20,000 people were tried on serious charges. (This assumption may be a bit low: in 1995, about 8% of all felony convictions were the result of verdicts, with the other 92% obtained by guilty pleas.) *See* BUREAU OF JUSTICE STATISTICS 1995, *supra* note 35, at 498 table 5.47. If we then assume that for less serious crimes only 1% of the cases result in a trial, again an assumption that might be too low, then there were an additional 60,000 trials. In other words, there were probably more than 86,000 criminal trials in the United States. Given these numbers, it is hard to imagine that the public was aware of much more than 1% of these trials (860).

[166] Some expressive utility, no doubt, arises from the message sent to the individual defendant. *See* Dan M. Kahan, *What Do Alternative Sanctions Mean?*, 63 U. CHI. L. REV. 591, 598 (1996) [hereinafter Kahan, *Alternative Sanctions*]. Furthermore, some expressive utility also arises from fact that there are *some* other individuals in a community who are aware of the trial's result. The point is not that there is no expressive utility, but rather that the expressive utility may be very limited.

[167] *See, e.g.*, Kahan, *Social Meaning*, *supra* note 160, at 615 ("What a community chooses to punish and how severely tells us what (or whom) it values and how much."). One might read into Kahan's thesis that society should approve of those actions that will lead to the validation of society's "authoritative moral values." *See* Kahan, *Secret Ambition*, *supra* note 160, at 422. However, this is a position that Kahan almost certainly would not share, given his conclusion that "[m]any of the social meanings that motivate citizens to support particular criminal law policies are *bad*." *Id.* at 499 (emphasis in original).

[168] Kahan believes that criminal law may increase social welfare by deterring crime and by "constructing *valued* social meanings." Kahan, *Social Meaning*, *supra* note 160, at 617 (emphasis added). So presumably, if we have a way of deciding which social meanings are *valuable*, we would be able to judge the value of enforcing (or not enforcing) various

crime to crime and from case to case.[169] The result is that expressive theories are not helpful in assigning values to our utility functions.

D. Retribution and Utilities

We have seen so far that neither deterrence nor expressive theories of criminal law seem likely to provide us, at least at present, with much guidance about the proper normative values for the various utility functions. But what about retributive theories, perhaps the dominant theory of criminal law at present?[170] Retributivists believe that the criminal should be punished because he deserves punishment.[171] At first blush, retribution seems no more likely to generate a consistent standard of proof than the theories discussed so far. In a retributive scheme, accurate convictions might provide utility because they at least provide the opportunity for retribution, while accurate acquittals provide at worst no utility because they avoid punishing someone who should not be punished.[172] Erroneous convictions would create costs by punishing someone who should not be punished, and erroneous acquittals create disutility by allowing those who should be punished to escape punishment. The amount of utility or disutility present for each possibility, though, depends upon the optimal punishment and, in a retributive model, this should depend upon the moral blameworthiness of the crime and the criminal.[173] In our criminal justice system, actual

---

criminal laws. But Kahan does not provide a normative guide as to which social meanings are valuable and which are not valuable.

[169] As we will see, the actual values that society should and does place on various verdict choices can vary significantly depending upon the particulars of the criminal case. *See infra* Part IV.

[170] *See* David Dolinko, *Retributivism, Consequentialism and the Intrinsic Goodness of Punishment*, 16 L. & PHIL. 507, 507 (1997).

[171] *See* MICHAEL S. MOORE, PLACING BLAME: A GENERAL THEORY OF THE CRIMINAL LAW 153 (1997) [hereinafter MOORE, PLACING BLAME]; MICHAEL S. MOORE, *The Moral Worth of Retribution*, *in* RESPONSIBILITY, CHARACTER, AND THE EMOTIONS: NEW ESSAYS IN MORAL PSYCHOLOGY, 179, 180 (Ferdinand Schoeman ed., 1987) [hereinafter Moore, *Moral Worth*]. For a recent discussion and summary of retributivist thinking, see Kaplow & Shavell, *supra* note 157, at 1228-49.

[172] The analysis in this paragraph is consequentialist, and retributivism is often (but not always) characterized as a non-consequentialist theory. *See* Dolinko, *supra* note 170, at 507-08. To the extent that some retributivists are consequentialists, this paragraph explains why retribution cannot lead to a high fixed standard of proof. To the extent that retribution is a purely deontological position, see the following two paragraphs. In the rest of this section, I will address a third alternative, the "mixed theory" approach.

[173] *See* J.L. Mackie, *Retributivism: A Test Case for Ethical Objectivity*, *in* PHILOSOPHY OF LAW 677, 679 (Joel Feinberg & Hyman Gross eds., 1991) [hereinafter Mackie, *Retributivism*] (noting that retributivism carries quantitative constraint).

punishments are likely to vary from whatever retributivists would believe to be optimal. Most obviously, the presence of mandatory minimums for certain crimes ensures that there will be at least some instances where a criminal is sentenced to a term that is longer than would be justified in a retributive approach. Although less draconian, determinate sentencing schemes such as the Federal Sentencing Guidelines have the same problem. Despite the large number of factors that a judge is required to consider, there are sure to be cases where a criminal is sentenced to a longer (or shorter) sentence than retribution would justify.

A retributivist might object that no amount of utility or disutility can be attached to an erroneous conviction, and that the avoidance of the harm that flows from an inaccurate conviction is simply an independent norm with which we must comply.[174] This can be seen as an incommensurability objection: that any consideration of a trade-off between erroneous convictions and the other possibilities would degrade or depreciate the "higher" value inherent in avoiding convictions of the innocent.[175] This claim suggests that there is something fundamentally unfair or unjust about convicting an innocent person of a crime that he did not commit. Thus, there cannot be (and perhaps should not be) any trade-off between inaccurate convictions and the other possibilities, a view that the Supreme Court has appeared to endorse at times.[176] This claim would appear to lead to the possible conclusion that the standard of proof cannot be calculated by purely consequentialist trade-offs of the sort suggested by Equation 3, and other means must be employed.

Such objections are usually focused on the inability to compare, say, expressive and interpretive dimensions against material and quantitative dimensions.[177] The claim here would be that there is simply no way of comparing the values of erroneously convicting the innocent and

---

[174] *See* Dolinko, *supra* note 170, at 510-11.

[175] *See* Richard H. Pildes & Elizabeth S. Anderson, *Slinging Arrows at Democracy: Social Choice Theory, Value Pluralism, and Democratic Politics*, 90 COLUM. L. REV. 2121, 2146-47 (1990) (describing hierarchical incommensurability).

[176] *See, e.g., In re* Winship, 397 U.S. 358, 364 (1970) ("When one party has at stake an interest of transcending value — as a criminal defendant his liberty — [the] margin of error is reduced as to him by the process of placing on the other party the burden of... persuading the factfinder at the conclusion of the trial of his guilt beyond a reasonable doubt." (*quoting* Speisel v. Randall, 357 U.S. 513, 525-26 (1958))).

[177] *See, e.g.,* Richard H. Pildes & Cass R. Sunstein, *Reinventing the Regulatory State*, 62 U. CHI. L. REV. 1, 70 (1995) ("CBA deals with the material or quantitative dimensions, not the interpretive and expressive ones."). For a broader description of possible interpretations of incommensurability, see Matthew Adler, *Law and Incommensurability: Introduction*, 146 U. PA. L. REV. 1169, 1170 (1998).

erroneously acquitting the guilty; the disvalue of erroneous convictions is distinctly higher than the disutility of erroneous acquittals.[178] This does not mean that it is true that the harms from an erroneous acquittal cannot be compared to those of erroneous convictions. Indeed, the existence of the traditional theory, which is premised upon Blackstone's maxim or some variant, suggests that we have long been comfortable with precisely such a trade-off.[179] And the Supreme Court has been willing to acknowledge such a trade-off in an analogous context of deciding the constitutional minimum size of a jury.[180] Acknowledging and evaluating the harm of letting the guilty go free has not degraded our moral commitment to protecting the innocent.[181] Indeed, it might be imagined that our willingness to discuss this trade-off has increased, not decreased, our respect for these values.[182]

Alternatively, a retributivist might well agree that such choices need to be made, but still contend that acquitting the innocent is a constraint on achieving the general aim of increasing the common good: a view I will refer to as negative retributivism.[183] One approach a negative retributivist might take is to say that the goal instead should be to require jurors have as close to absolute certainty as possible.[184] The difficulty with such a claim is that it would be theoretically impossible to require absolute certainty (or something quite close to it). For instance, the criminal justice system, if it valued primarily the avoidance of

---

[178] *See* Pildes & Anderson, *supra* note 175, at 2147.

[179] *See supra* notes 41-43 and accompanying text.

[180] *See* Ballew v. Georgia, 435 U.S. 223, 234 (1978) (relying upon study that set jury size based on relative risks of erroneous acquittal and erroneous convictions).

[181] *See infra* text accompanying notes 331-333 for an argument on how the existing reasonable doubt instruction in fact facilitates this process.

[182] A purely negative retributive account might suggest that even though this trade-off has had positive results, it is still impermissible because inaccurate convictions are unique in their harm. Such an account would by necessity be objective, because it seems clear that if negative retributivism is viewed subjectively, then our subjective moral reasoning has led us to engage in this trade-off; the persistence of Blackstone's maxim and its variants in our justifications of the criminal law suggests that we are not committed to a pure subjective negative retributive claim. As for an objective theory, there seems little reason to believe that objective negative retributive accounts are distinct from objective general retributive accounts.

[183] *See* H.L.A. HART, PUNISHMENT AND RESPONSIBILITY 8-11 (1968). A negative retributivist view is not a justification for punishment, but rather a justification against punishment. *See* Mackie, *Retributivism*, *supra* note 173, at 679.

[184] *See, e.g.*, Laurence H. Tribe, *An Ounce of Detention: Preventative Justice in the World of John Mitchell*, 56 VA. L. REV. 371, 388 (1970) [hereinafter Tribe, *An Ounce of Detention*] ("Thus, guilt beyond a reasonable doubt represents . . . a standard that seeks to come as close to certainty as human knowledge allows. . .").

convicting the innocent, might require convictions only in cases where the defendant has confessed.[185] But there is no such requirement or really any other mechanisms — other than the reasonable doubt instruction — to ensure a high level of certainty prior to a conviction. The failure of the system to require near certainty is an implicit acknowledgment that other concerns of fairness and justice — for the victims and for the other members of society — must somehow trump our moral commitment to the defendant.

Thus, it is far more plausible that a negative retributivist might simply contend that the standard of proof can be set by a straight forward acknowledgment that the costs of erroneous convictions are greater than the costs of erroneous acquittals. Of course, this still leave open the difficulty of valuing the benefits from accurate convictions. Even assuming that the negative retributivist could safely ignore these benefits, though, she is left with the task of justifying a particular ratio of costs. Given the widely different sets of estimates that have been given by commentators, this seems like a hopeless task.[186] Finally, even if negative retributivists could agree on a single ratio of costs of erroneous verdicts, they must also find a way of justifying the conclusion that the *same* ratio should apply to all cases: from capital murder cases to traffic infraction cases.

The result is that retributive theory is also unlikely to lead to a single standard of proof for all cases. The failure of the retributive theory to require a single, high standard of proof in all cases does not mean that the theory is irrelevant to understanding reasonable doubt. Indeed, the strong moral sentiment of society that supports the negative retributive account possibly explains why we disvalue (in general) inaccurate convictions so much more than inaccurate acquittals (particularly compared in civil cases).[187] It may well be that the shame we associate with convictions leads to an understanding that the moral harm from an erroneous conviction is so much greater than the moral harm that attaches to an erroneous finding of liability in a civil case.

---

[185] This would not actually guarantee certainty, because some defendants would confess falsely, either voluntarily or involuntarily. Nonetheless, if such a confession was a necessary condition of conviction, it would serve to lessen the number of inaccurate convictions.

[186] *See supra* text accompanying notes 61-68.

[187] *See* Mackie, *Retributivism*, *supra* note 173, at 781-82 (noting strength of this sentiment).

E. Empirically Derived Utilities

The result is unlikely to be any different under other theories of criminal law.[188] Indeed, it appears quite likely that no single theory can properly explain our use of criminal law.[189] Instead, the application of criminal law is in some instances the result of one theory, and in other instances the result of a different theory. The consequence is that there is even less reason to believe that we will be able to ever generate a standard of proof of a fixed level of certainty based on existing models.

A way around all of these objections might be to attempt to derive the values that society itself attaches to each utility or disutility, and then calculate the standard of proof based on these. Conveniently, some researchers have actually attempted to measure the values that their research subjects attached to various verdict outcomes under different conditions. For instance, Stuart Nagel had his subjects (students at the University of Illinois) rank the undesirability of the erroneous verdicts from –100 to 0 or the desirability and undesirability of all four possible outcomes between +100 and -100.[190] Taking these values and plugging them into the decision theory model results in a standard of proof set at a .55 level of certainty.[191] Nagel also reports that there were differences in the standard of proof that was calculated based on each individual subject's stated values for the utilities.[192] Although he does not report how great the actual differences are, it is likely that they were large, given the great differences he reports in subjects' responses when asked to rate the disutility of acquitting a guilty defendant.[193] Other researchers

---

[188] The most important theory left is rehabilitation. Just as with deterrence and retribution, the value of rehabilitation is likely to vary under the circumstances of the particular case.

[189] *See* Kaplow & Shavell, *supra* note 157, at 1246 n.697 (noting current confusion about justifications for criminal sanctions among academics). In a related vein, Kahan has suggested more generally that it might not be appropriate "for a liberal society to concern itself with the quality of its citizens' values." Kahan, *Alternative Sanctions*, *supra* note 166, at 630.

[190] *See* Nagel et al., *supra* note 56, at 360-62 & fig.2.

[191] Nagel, *supra* note 74, at 192 & n.5.

[192] *Id.*

[193] *Id.* at 191. These wide variations are consistent with the wide variations that are reported when subjects are directly asked their standard of proof. *See* Power, *supra* note 9, at 105-06 (citing C.M.A. McCauliff, *Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?*, 32 VAND. L. REV. 1293, 1295-96 (1982)).

Nagel's methodology consisted of placing a value of –100 on the most undesirable of the possible outcomes (most subjects responded that this was an erroneous conviction) and then asking the subjects the position of the other undesirable outcome (an erroneous acquittal) on the –100 to 0 scale. Nagel, *supra* note 74, at 191. Nagel reports that the results

have reported similar results.[194]

This data suggests that using the values that society places on each outcome would lead to a much lower standard of proof than the traditional view suggests. Of course, there are several reasons why this data might not be fully persuasive.[195] First, while the values may hold for the individual case to which the subjects were exposed, it is not clear how well these values would hold up over all cases.[196] But this just suggests that society may have different utility values in different cases, with a different standard of proof in different cases (a possibility I will defend in Part IV).[197] The only way to avoid this would be to find a way

---

here range from –5 to –95. *Id.* If we were to use Kaplan's and Cullison's original equation (Equation 2; $SP = 1/(1 + U_{ag}/U_{ci})$, this would generate results ranging from .95 to .51. Obviously, this is a wide variance in the possible standards of proof.

[194] Francis Dane asked subjects their values both before and after deliberation — prior to deliberation, but after hearing the trial, the resulting standard of proof was .524; after deliberation, the standard of proof actually dropped to .516. Francis C. Dane, *In Search of Reasonable Doubt: A Systematic Examination of Selected Quantification Approaches*, 9 LAW & HUM. BEHAV. 141, 152 (1985). (Dane clearly asked for utility values for all four possibilities: accurate acquittals and convictions as well as erroneous acquittals and convictions. *Id.*) Similarly, Reid Hastie, using data from a study by William Thompson and his colleagues determined that jurors who are "death-qualified" (i.e., eligible to sit on a jury in a case where the government will ask for the death penalty) will set *SP* at .50 and the "death-excludable" jurors will place it somewhat higher at .58. Hastie, *supra* note 22, at 105 table 4.3. Hastie is not clear about what he does with certain data points, such as the regret of a subject when she voted to convict a defendant of manslaughter, but the defendant was actually guilty of second-degree murder.

The original study is in Thompson et al., *supra* note 135. Thompson and his colleagues did not directly calculate *SP* because they believe that data on the disutility of various verdict choices "probably are not an accurate index of the *absolute* level of jurors' thresholds of conviction." *Id.* at 109. They do believe, however, that it does provide a measure of the relative levels of jurors' thresholds of conviction. *Id.*

[195] Other data that Hastie has collected for his meta-analysis is generally consistent with that described in the text. *See* Hastie, *supra* note 22, at 105 table 4.3.

[196] For instance, Nagel reports that while the standard of proof for a robbery and a rape case were the same when balanced across all subjects, breaking down the responses by gender revealed that women had a significantly lower standard of proof in the rape case than in the robbery case and that men had a higher standard of proof in the rape case than in the robbery case. Nagel et al., *supra* note 56, at 369-70.

[197] Similarly, there is reason to believe that the knowledge about the punishment attached to the crime might alter the ordering of the values. For instance, Nagel notes evidence that capital punishment may increase the decision criterion. Nagel et al., *supra* note 56, at 369 n.15. However, there is also reason to believe that the severity of the punishment has no actual effect on jurors' decision to convict or acquit where capital punishment is not implicated. *See* Jonathan L. Freedman et al., *Severity of Penalty, Seriousness of the Charge and Mock Jurors' Verdicts*, 18 LAW & HUM. BEHAV. 189, 199 (1994). *But see* Martin F. Kaplan, *Setting the Record Straight (Again) on Severity of Penalty: A Comment on Freedman et al.*, 18 LAW & HUM. BEHAV. 697, 697-99 (1994) (criticizing this study). This may make rough sense. Capital punishment may vastly increase the costs of erroneously convicting the innocent without corresponding benefits. Increases in

of weighing the various cases and coming to an "average" standard of proof to be applied to all cases. But even assuming that such an averaging would make sense (something I dispute in Part IV), the resulting average is still likely to be lower than the level of certainty we commonly associate with proof beyond a reasonable doubt.

A second objection would be that these surveys may not yield accurate values. A foundational question has to be: in what context should society be asked to assign values to the possible outcomes and how much information should society have before being asked their values? Should it be done in the abstract or only in the context of a real case? The surveys done by these researchers were either done in the abstract or in the context of a hypothetical case that was presented to the subjects; as a result, the research may not reflect the values of the subjects when faced with a real case.[198] This distinction only matters, however, if society, after hearing a real case, would tend to assign values that result in a higher standard of proof. And assuming that the decisions of jurors are a good reflection of the decisions society would make, then as the discussion in Part II revealed, there is good reason to believe that even in the context of a real case, society is likely to assign a value lower than we commonly assume.

This objection could be extended, though, to suggest that society (as represented here by the subjects) would reach different results if it had more information about the possible repercussions of its decisions. For instance, subjects might alter the number they assign to a value if they knew how the number was translated into a standard of proof. Indeed, Nagel reports that some subjects revaluate their stated utilities when they were made aware that they were generating low standards of proof.[199] But even if more information would tend to increase the

---

imprisonment terms alone, without other benefits resulting from imprisonment, may create offsetting shifts in the utility functions. *See also* BENNETT & FELDMAN, *supra* note 19, at 160.

[198] *See* Nagel et al., *supra* note 56, at 367; Dane, *supra* note 194, at 146-47. For a summary of some criticisms of jury research, see Elwork et al., *supra* note 103, at 50-56.

[199] Nagel et al., *supra* note 56, at 368. Other information, such as the way in which the jurors are instructed on the standard of proof, might also have an effect on the way in which subjects order the various utilities. *Id.* at 376-81. Most interestingly, Nagel's research shows that when the instruction explicitly quantifies the reasonable doubt standard, subjects did alter their values for the utilities such that the level of certainty required by the reasonable doubt standard increased (although generally not so high as to actually meet the level of certainty that the instruction itself demanded). *Id.* at 378 table I; *see also* Dorothy K. Kagehiro & W. Clark Stanton, *Legal vs. Quantified Definitions of Standard of Proof*, 9 LAW & HUMAN BEHAV. 159 (1985) (noting that experiments demonstrated that subjects applied to their analyses definitions' difficulty level). But it's hard to see why the fact that the instruction varies the standard of proof matters in this context when we are

standard of proof for *some* subjects, it does not appear to do so enough *for all subjects* that the resulting standard of proof would be at anything like .90, particularly given the very low numbers that are reported by Nagel and others.[200]

So, despite our articulated preference for a fixed, high level of certainty before convicting in criminal cases overall, the traditional view cannot provide a normative justification for such a result. The model fails because the existing justifications for the substantive criminal law cannot lead to a fixed decision standard of proof. The alternative, relying on values assigned by "society" for the utility functions, appears to lead to a much lower standard of proof than is conventionally assumed, and perhaps away from a fixed standard of proof.

IV. VARYING REASONABLE DOUBT

The lesson of the first three Parts of this Article is that there is a significant disconnect between the ideology that underlies the traditional theory and the realities of our criminal justice system. The traditional view posits that because it is much worse to convict an innocent person, we do and/or should require a high level of certainty before convicting a defendant at trial. But as Part II shows, jurors do not always require such a high level of certainty, and, as Part III demonstrates, existing variants of the traditional view cannot justify a high level of certainty in all criminal cases.

The puzzle then is whether the descriptive reality can be meshed with a normative explanation of reasonable doubt. Other commentators have suggested the answer is no. These commentators, after noting the possible divergence between the ideological commitment to a high level of certainty and what jurors actually do, have concluded that the instruction must be fixed or that other steps must be taken to ensure that a high level of certainty is required of decision makers in criminal

---

trying to determine the level of certainty the instruction itself should be generating. To the extent instructions are relevant, they are means by which we achieve a result, not independent justifications for a result.

[200] We might also question whether surveys of individual opinions about the various utilities tell us what we would want to know in a representative democracy. Even assuming that the survey is accurate, *who* should we be surveying: citizens, registered voters, likely voters? Surveys of the entire population or even of some subset of the population might not end up generating the data we actually would want. Furthermore, surveys provide no accounting for strength of preferences, since each person's view is accorded equal weight. But there is no specific reason to believe that altering the population subset from which the subjects are drawn would result in a large enough change to significantly alter the results of the data.

cases.[201]

My purpose here is to challenge that conclusion and to suggest that a flexible reasonable doubt standard is preferable to a standard that would require a single, fixed level of certainty. As I stated in Part III, there is (and probably can be) no agreement about the proper ordering of utilities across all cases, so no single level of certainty should govern all cases. Part IV.A demonstrates that, despite this absence of system-wide agreement, it is predictable that society will achieve some consensus about the proper ordering of utilities in particular cases, at least relative to other cases. This implies that it makes sense that the standard of proof ought to be higher in some cases than in others. For instance, in death penalty cases the standard of proof perhaps should be even higher than what we commonly associate with "proof beyond a reasonable doubt." In some other cases, the preferable amount of certainty will be less than our ideology suggests.

Part IV.B then shows that, even though there are some predictable error costs from a flexible reasonable doubt standard, it is likely that a fixed standard of proof would incur even greater error costs. This conclusion is buttressed by the fact that it is the jury who decides the applicable standard. Jurors, left to decide the standard of proof for themselves in any case, are likely to achieve results that are preferable.

Finally, Part IV.C discusses the objections of some commentators to engaging in this sort of weighing of costs and benefits in setting the criminal standard of proof. These commentators claim that there is an alternative explanation for the reasonable doubt standard: that it promotes the legitimacy of the criminal justice system. Part IV.C concludes that such explanations are unsatisfying, because the alleged benefits of legitimacy, upon closer examination, appear to be illusory.

A. The Social Value of Variable Reasonable Doubt

Return again to Kaplan and Cullison's model. Part III above demonstrated that this model, when framed as a normative theory, had two major problems. First, the model had to account not just for erroneous convictions and acquittals, but also accurate convictions and

---

[201] *See supra* note 9 and accompanying text. *But see* Shaviro, *supra* note 11, at 543 n.43 (not taking position on desirability of requiring different levels of certainty in different cases); Elisabeth Stoffelmayr & Shari Seidman Diamond, *The Conflict Between Precision and Flexibility in Explaining "Beyond a Reasonable Doubt,"* 6 PSYCHOL. PUB. POL'Y & L. 769, 781 (2000) (stating that flexibility may be "justifiable"); *but see also* Posner, *Economic Approach to the Law of Evidence*, *supra* note 23, at 1506 (suggesting that lower standard of proof may be acceptable because few errors are probably made).

acquittals. Second, there presently exists no clear way to determine the various utility functions for all four variables, a fact brought home by the absence of any consensus over the correct variables in the literature.[202] The result of these problems is that we cannot, with any true confidence, generate a normatively preferable standard of proof to apply to all cases.

This does not mean that the model cannot be used to aid in decisions as to the proper standard of proof in particular cases. The model may predict that we should prefer a higher standard of proof in some cases relative to other cases. Of course, doing so would require some consensus, even if only a rough one, over what the proper utilities are for the possible verdicts in that case. I have suggested that obtaining a consensus over the possible utilities for all criminal cases is a hopeless task: that we are too split to come to a consensus that would result in a single set of values that would apply in all cases.[203] But setting the relative utilities across classes of cases might not be as difficult. For instance, we may be able to (and, I believe, can) conclude that in some classes of cases, relative to other classes of cases, the utility of a decision to convict is sufficiently lessened (or the utility of a decision to acquit is sufficiently heightened) that the preferred standard of proof in the one set of cases should be higher than in others.[204]

This sort of thinking has long been inherent in the literature about the death penalty. Judge May, in his influential article on the reasonable doubt standard in 1876, noted that the standard began to take shape at a

---

[202] Compare, for instance, the values discussed above that were given by Tribe, Milanich and myself. *See supra* notes 62-66 and accompanying text. Although each author (other than myself, perhaps) has given a compelling rationale for his or her own ordering of the utilities, there seems to be little basis on which to select amongst them.

[203] *See supra* text accompanying notes 146-149.

[204] *See* Kaplan, *Decision Theory and the Factfinding Process*, *supra* note 52, at 1073-74 (observing that it is likely that standard of proof in his model will vary depending on particular case, but making no attempt to see if this in fact occurs); Stoffelmayr & Diamond, *supra* note 201, at 781 (discussing possibility that flexibility may result from sentence severity); *see also* Risinger, *supra* note 20, at 442-45 (suggesting that Wigmore may have been up to something similar in arguing that prosecution in a criminal case should be permitted to refuse defendant's offer to make a judicial admission as to certain issues); Nance, *Evidential Completeness*, *supra* note 52, at 624 (suggesting that there may be various limited ranges within which all standards of proof may vary). Dale Nance has recently suggested that jurors are given limited flexibility which they use to avoid the conjunction effect, referenced in note 23, *supra*. *See* Nance, *Evidential Completeness*, *supra* note 52, at 1574 n.77. This Article suggests that this is at most a second best explanation. First, as detailed in Part II.B.3, there is little reason to believe that jurors pay attention to jury instructions in general, let alone any specific instruction to weigh each element separately. Second, the universal availability of flexibility in criminal cases would seem to be far overbroad for the few limited number of cases in which there truly are two (or more) issues for the jury to decide and that therefore implicate the conjunction effect.

time when

> the penal code of England was a fearfully bloody code. Death, without benefit of clergy, was denounced against a multitude of misdoings which would now be considered, if offences at all, offences of a comparatively trivial character. The consequences of conviction to the unfortunate prisoner were not only fearful, but they were irremediable.[205]

This was consistent with the thinking of other writers at the time.[206] Stated in decision theory terms, Judge May recognized that, in a criminal justice system in which the death penalty was common, the disutility of erroneous convictions might be greater, resulting in a lowering of the overall utility of a decision to convict, which would lead in turn to an increase in the amount of certainty that a jury should require before reaching a decision to convict.[207] Put more bluntly: society may want a higher degree of certainty of guilt before sentencing someone to death than in other, non-capital cases.

Decision theory suggests that Judge May's intuition is correct in the event that there are cases where the disutility of erroneous capital verdicts is higher than the disutility of erroneous noncapital verdicts. Start with the easiest case and assume that the only difference between a capital case and a noncapital case is the relative amount of disutility associated with an erroneous conviction. If this is so, then the standard of proof should be higher in the capital case. For example, if in a typical criminal case we believe that the proper ordering of the relative utilities is $U_{cg} = 1$, $U_{ai} = 0$, $U_{ag} = -1$ and $U_{ci} = -10$, this results in $SP = .83$.[208] If we then assume that in a capital case the disutility of an erroneous

---

[205] May, *supra* note 40, at 651-52; *see also id.* at 659-60 (criticizing use of reasonable doubt standard when defendant is not charged with serious crime).

[206] *See* Thomas A. Green, *A Retrospective on the Criminal Trial Jury, 1200-1800, in* TWELVE GOOD MEN AND TRUE, 398 (J.S. Cockburn & Thomas A. Green eds., 1988) ("As self-protective and sincerely believed rationalizations of its own position and authority, the state emphasized what had long been mainly the case: that conviction at a capital level required virtually absolute proof."); GREEN, *supra* note 123, at 286 ("For lay writers especially, the most important factor in acquittals and partial verdicts was the very high threshold of proof that the court required for conviction of a capital offense.").

[207] Indeed, Judge May noted that Hale's and Blackstone's comments about the proper ratio of errors in criminal cases were directly linked to those authors' concerns with the harshness of the sanctions under the English penal code. *See* May, *supra* note 40, at 653-54. Kaplan also implicitly understood that the use of the death penalty might tend to raise the standard of proof for first-degree murder. *See* Kaplan, *Decision Theory and the Factfinding Process*, *supra* note 52, at 1074-75; *see also* NAGEL & NEEF, *supra* note 59, at 208 n.16 (noting that jurors demand higher probability of guilt in capital cases).

[208] *See supra* text accompanying note 66.

conviction is five times that in the noncapital case (with the result that U_{ci} = -50), the result is that *SP* = .96, a higher value. Furthermore, this result is not dependent upon the particular set of values used here. So long as one places greater disvalue on erroneous convictions in capital cases than in noncapital cases, the standard of proof should be higher in a capital case.

The conclusion that the standard of proof should vary in capital cases still holds even if we drop the assumption that the only difference between capital and noncapital cases is the disutility of an erroneous conviction. Instead, assume that not only is the utility of an erroneous conviction changed, but so too are the utility of accurate convictions and the disutility of an erroneous acquittal.[209] Even if changes to these other utilities are factored in, the standard of proof should still change. To prevent an increase in the standard of proof, the rise in the combined disutility of erroneous acquittals and utility of accurate convictions would have to be proportionate to the rise in the disutility of an erroneous conviction. For purposes of exposition, I have assumed that the disutility of an erroneous conviction in a capital case is five times greater than in a noncapital case, so the combined utility of accurate convictions and disutility of erroneous acquittals would have to be five times greater as well. If so, we would get U_{cg} = 5, U_{ai} = 0, U_{ag} = -5 and U_{ci} = -50, which would result in *SP* remaining at .83. But if we assume instead that while the disutility of an erroneous conviction is five times greater, but the utility of an accurate conviction and the disutility of an erroneous acquittal are only two times greater, we get *SP* =.93.

So while it is hypothetically possible that the use of the capital sanction will not result in a change in the standard of proof, achieving that result is dependent upon what seems to be an extremely dubious assumption: that in capital cases, the proportionate rise in the utilities and disutilities for accurate convictions and erroneous acquittals would

---

[209] Accurate convictions are more valuable when the penalty is death because society can be sure that the defendant will never offend again, while there is even more disutility from an erroneous acquittal because only the worst offenders are likely to be charged with a capital crime and they may be disproportionately likely to offend again and do the worst kinds of things. In addition, given the greater publicity that comes with capital cases (at least, on average, in comparison to noncapital cases), the acquittal of guilty defendants will lead to decreased faith in the criminal justice system and an increased feeling that "crime pays," at least if there is a widespread belief that the defendant was in fact guilty. Of course, as to this last objection, one could note that there is an offsetting disutility in capital cases from erroneous convictions: because capital cases obtain more publicity, an erroneous conviction (to the extent that outsiders know that the conviction is erroneous) also decreases faith in the criminal justice system and a feeling, at least among some groups in society, that there is little point in obeying the law.

be exactly the same. In fact, the situation might even be more complicated than I have described it thus far, because I have assumed throughout that the utility of an accurate acquittal is zero. If we instead assume that there is a positive utility associated with accurate acquittals (as Tribe and Milanich do[210]) the effects of the use of the death penalty on the standard of proof are further complicated, because we have to account for four variables, not three. For instance, if $U_{ai} = 1$, then the standard of proof in a capital case would remain the same if and only if the amounts of change of utilities and disutilities of the other three values were set to precisely offset the effects of the rise for disutility of erroneous convictions.[211] This result can be generalized. What is always essential to ensure that the standard of proof remains unchanged, despite a change in the values of the utility functions, is that the value of the fraction remain unchanged. Thus, in the example I have previously used, where $U_{cg} = 1$, $U_{ai} = 0$, $U_{ag} = -1$ and $U_{ci} = -10$, the standard of proof will change in a capital case unless the value of the fraction remains 1/5. If the values change in any way that results in different values for the fraction, then the value of *SP* must change.[212] As a result, we can indeed theorize that the standard of proof should rise where the death penalty is implicated. And this result is consistent with the empirical results of at least one set of researchers.[213]

Death penalty cases are not the only ones in which we should see a change in the standard of proof that society would prefer.[214] Indeed, it

---

[210] *See supra* notes 62-64 and accompanying text.

[211] Therefore, if in noncapital cases, $U_{cg} = 1$, $U_{ai} = 1$, $U_{ag} = -1$ and $U_{ci} = -10$, then *SP* would be .85. If in capital cases U_{ci} rises to –50, then one (but not the only) way of ensuring that *SP* = .85 would be to create similar proportionate changes in the other values, such that $U_{cg} = 5$, $U_{ai} = 5$, $U_{ag} = -5$. The particular values do not matter; what is essential is that for the standard of proof to be unchanged in this example, the value of our fraction $(U_{cg} - U_{ag})/(U_{ai} - U_{ci})$ must remain 2/11.

[212] Of course, Milanich and Tribe both assume that erroneous convictions have no value: $U_{ci} = 0$. Presumably, both of them would believe that there is even less value for convictions in capital cases then in noncapital cases. Because what is important is not the absolute values of these utilities, but rather their *relative* values, *see* Part I.C, *supra*. Milanich and Tribe could adjust for this simply by increasing the relative values of the other verdict possibilities. This would not change the fact that the only way to assume no change in the standard of proof in capital cases is to assume that the value of the fraction $(U_{cg} - U_{ag})/(U_{ai} - U_{ci})$ remains unchanged.

[213] *See* Nagel et al., *supra* note 56, at 369 n.15.

[214] The qualification "that society would prefer" is important. In practice, death penalty jurors are frequently skewed by the practice, permissible under *Lockhart v. McCree*, 476 U.S. 162 (1986), of disqualifying a potential juror if the prospect of the death penalty would affect her deliberation. *See* Witherspoon v. Illinois, 391 U.S. 510 (1968). Of course, what the preceding discussion illustrates is that it would be odd if the prospect of the death penalty

seems reasonable to assume that there are whole classes of cases where the values of the various utilities are sufficiently different that we can conclude that the standard of proof should be higher in one case than in the other. An example may help further illustrate the point. Assume two defendants, one charged with theft from inside an automobile, the other with rape of a child. Assume that in both cases, the only real issue is identification. Also assume that in both cases, the defendant has been convicted of this crime on at least one previous occasion. In this example, there will be two primary reasons that the burden of persuasion should vary: the characteristics of the crime and the characteristics of the offender. I will discuss their combined effect before turning to how they might act independently.

First consider the auto theft case. A uniformed police officer testifies to having seen an individual who approached a parked car and broke one of the front windows of the car. After reaching into the car, the individual began to walk away. The officer — who was a significant distance away — immediately radioed defendant's description and location to police officers also patrolling the neighborhood. The officer then gave chase, but during the chase, lost sight of the individual. Based on the officer's description, however, other officers apprehended defendant and, upon his arrest, recovered bridge tokens, as well as some loose change. At trial, the owner of the vehicle testifies that the ashtray of the car had contained change and bridge tokens (the officer testified the ashtray was empty when he returned to the car). The defendant denies that he was the individual seen by the officer and that he had been walking from a nearby bodega. In its cross-examination, the prosecution elicits testimony from the defendant that he has been convicted of theft from an automobile on two prior occasions.[215]

---

did not affect the juror's deliberation. And, indeed, research appears to indicate that "death-qualified" jurors have a lower standard of proof relative to "excludable" jurors. *See* HANS & VIDMAR, JUDGING THE JURY, *supra* note 101, at 233; Thompson et al., *supra* note 135, at 108; *see also* Phoebe C. Ellsworth, *Some Steps Between Attitudes & Verdicts*, *in* INSIDE THE JUROR, *supra* note 20, at 42, 56-57 (describing different levels of expression of regret for erroneous convictions and acquittals among death-qualified and excludable jurors in another article on same research). Kaplan notes that the use of death-qualified jurors may be the only way in which the standard of proof is low enough that the criminal justice system is able to obtain guilty verdicts in capital cases. *See* Kaplan, *Decision Theory and the Factfinding Process*, *supra* note 52, at 1075. I doubt that expanding the juror pool (by overruling *Witherspoon* and *Lockhart*) would eliminate guilty verdicts, but I do believe it would lower the number of them because the standard of proof would rise.

[215] Assume that the evidence is admissible as similar acts evidence under FED. R. EVID. 404(b) as evidence of identity. *See, e.g.*, Howell v. State, 627 So. 2d 1134, 1138 (Ala. Crim. App. 1993) (affirming admission of evidence of prior burglary conviction to show identity

Second, consider the rape case. A ten-year old girl is abducted. The assailant drives the girl to a secluded area and rapes her. The assailant eventually releases the girl and the parents take the girl to the hospital, where a physician treats the girl for her injuries. The physician uses a rape kit to collect evidence of the attack. Nine days later, the girl identifies the defendant in a photographic array. In the meantime, a police criminologist examines the sexual assault kit. He testifies at trial that he followed standard department procedure, which was to determine whether there had been sexual contact. (Blood group testing, however, was not performed because the samples were too small and DNA tests were not available.) At trial, the defendant's principal defense is that the girl has misidentified him. Both defendant and others testified that his car did not fully match the girl's description. In addition, a witness testified that the defendant had been with her on the night of the assault. At trial, the prosecution elicits testimony from the defendant that he has been convicted of sexual assault of a minor on a prior occasion.[216]

What are the utilities and disutilities in these two cases? Consider the costs and benefits of a decision to convict: for this decision, the utility consists of the potential benefit derived from an accurate guilty verdict balanced against the potential cost of an erroneous conviction. In both of our hypothetical cases there are several potential benefits from an accurate conviction. We punish someone who is engaged in criminal behavior, thereby incapacitating him from such activity for at least some time. We may create deterrence against other potential offenders (thieves or rapists) from engaging in such behavior because they become aware that there is some likelihood that they will be punished. We may also discourage such activity because we have expressed society's disapproval of such activity.[217] But there are also offsetting costs of an

---

of offender in subsequent burglary prosecution); Carter v. State, 406 S.E.2d 238, 239 (Ga. Ct. App. 1991) (affirming admission of prior theft conviction in subsequent robbery prosecution to show identity, motive, plans, scheme, bent of mind, and course of conduct); Marks v. State, 357 S.E.2d 299, 299 (Ga. Ct. App. 1987) (affirming admission of prior theft conviction in subsequent theft prosecution to show identity); *see* JACK B. WEINSTEIN & MARGARET A. BERGER, WEINSTEIN'S FEDERAL EVIDENCE § 404.22[5][c] (Joseph M. McLaughlin ed., 2d ed. 2000). The evidence might also come in as impeachment evidence under FED. R. EVID. 609. *See, e.g.*, State v. Rendon, 715 P.2d 777, 781 (Ariz. Ct. App. 1986) (affirming admission of evidence of prior theft conviction in subsequent theft case under state version of Rule 609); People v. Graham, 327 N.E.2d 261, 265 (Ill. App. Ct. 1975) (same); State v. Pittman, 2001 WL 589162, at *5 (Tenn. Crim. App. May 31, 2001) (same).

[216] Here, the evidence is clearly admissible under FED. R. EVID. 414.

[217] We also may obtain retribution against the offender and/or allow for the rehabilitation of the offender.

erroneous conviction that we must consider. There is the loss to the defendant of his freedom even though he committed no crime. Deterrence may be undermined because the real defendant has gone unpunished (and thus undeterred) and other potential offenders may think that they can "get away" with this crime.[218] We may also undermine the benefits of expressing our disapproval of such activity by punishing someone who has not in fact committed the crime.

There are also possible differences between the two cases. For instance, an erroneous conviction creates losses because we are incapacitating a defendant who has committed no crime. This cost may be significantly different for our theft case and for our rape case. Remember, in both cases we know that the defendant has committed this crime before, but the cost of incapacitating a former thief and a former rapist, even when the defendant has not committed a subsequent crime, *could* be different.[219] Some states have passed laws providing for the civil commitment of violent sexual predators,[220] while other states have passed so-called "Megan's laws," providing for the registration of sexual offenders with local police.[221] These types of statutes suggest the possibility that, far from being a certainly bad thing, erroneously convicting a past rapist may have social benefits. These benefits may not *exceed* the costs, such that an erroneous conviction is a desirable outcome, but nonetheless an erroneous conviction may be less undesirable in some cases than in others.

It is certainly debatable whether civil commitment statutes and Megan's laws represent a normatively satisfying approach to the problem of sexual offenses.[222] And, indeed, just as there may be benefits

---

[218] *See also* Posner, *Economic Approach to Legal Procedure*, *supra* note 13, at 412 (noting that punishing innocent presumably reduces expected punishment costs of crime).

[219] *See, e.g.*, Alexander D. Brooks, *The Constitutionality & Morality of Civilly Committing Violent Sexual Predators*, 15 U. PUGET SOUND L. REV. 709 (1992) (arguing that Washington's Sexually Violent Predator statute is both constitutional and moral); John Kip Cornwell, *Protection & Treatment: The Permissible Civil Detention of Sexual Predators*, 53 WASH. & LEE L. REV. 1293 (1996) (arguing that indefinite civil commitment of sexual predators is constitutional); *see also* Christopher Slobogin et al., *A Prevention Model of Juvenile Justice: The Promise of* Kansas v. Hendricks *for Children*, 1999 WIS. L. REV. 185 (arguing that preventive model is appropriate for juvenile offenders who might offend again).

[220] KAN. STAT. ANN. §§ 59-29a01 to 59-29a20 (2001); WASH. REV. CODE §§ 71.09.010 to 71.09.902 (2002).

[221] N.J. STAT. ANN. §§ 2C:7-1 to 2C:7-11 (West 1995); 42 PA. CODE §§ 9791-99.6 (1998).

[222] Many commentators have criticized the use of a preventive detention scheme. *See, e.g.*, C. Peter Erlinder, *Minnesota's Gulag: Involuntary Treatment for the "Politically Ill"*, 19 WM. MITCHELL L. REV. 99 (1993); Gary Gleb, Comment, *Washington's Sexually Violent Predator Law: The Need to Bar Unreliable Psychiatric Predictions of Dangerousness from Civil Commitment Proceedings*, 39 UCLA L. REV. 213 (1991); John Q. La Fond, *Washington's*

from erroneously convicting past rapists, there may be some benefits from erroneously convicting past thieves. Nonetheless, the presence of the civil commitment statutes and Megan's laws, and the fact that they apply to sexual offenses and not to theft offenses, suggests that elected representatives of society (at least in those states that have approved such regimes) believe that there is a difference, one that may alter the utility calculus in such cases. I am not suggesting that, as a normative matter, such errors are preferable; to the contrary, I am assuming that we do not know if such distinctions can be normatively justified. It does seem clear, however, that at least some members of society have indicated that incapacitating sexual offenders may be justifiable, while such incapacitation of thieves may not be. The result is that society may view the disutility of an erroneous conviction as lower in our hypothetical rape case than in our hypothetical theft case.

Of course, we may also associate a greater utility with *accurately* convicting rapists rather than thieves. We can suppose that the length of the sentence for an accurate rape conviction here will exceed the likely sentence for a theft conviction, in part because society has more interest in deterring individual instances of rape than theft.[223] Furthermore, civil commitment statutes and Megan's laws generally require that the defendant be convicted at least once, so an accurate conviction has the additional utility of allowing society to limit the activities of past rapists, either through civil commitment or through posting of names in accordance with such laws. Thus, taken as a whole, we can conclude that society believes that the disutility associated with convicting the innocent may not be as great in our hypothetical rape case as it would be in our hypothetical theft case. But at the same time, the utility associated with an accurate conviction in our rape case may be higher as well.

The same sort of analysis applies to the decision to acquit in our two hypothetical cases. In both cases, there are benefits from an accurate acquittal: for example, ensuring accuracy in the system (which may increase respect for the legal system) and ensuring that an innocent person is not imprisoned. Similarly, there are costs from an erroneous

---

Sexually Violent Predators Statute: Law or Lottery? A Response to Professor Brooks, 15 U. PUGET SOUND L. REV. 755 (1992); Stephen J. Morse, *Blame & Danger: An Essay on Preventive Detention*, 76 B.U. L. REV. 113 (1996). For a summary of the criticisms to preventive detention schemes, *see* Slobogin et al., *supra* note 219, at 200-04.

[223] In fact, we might conclude that the difference in the sentences for rape and theft would *underestimate* the difference in the marginal costs to society of a particular commission of each offense. This might be true if we believed that the probability of a particular instance of theft being detected and punished was less than the probability of a particular instance of rape being detected and punished.

acquittal: such a decision undermines deterrence, as well as the expressive authority of the law, by allowing a criminal actor to go free.

There may also be differences between the rape and theft cases as to the utility of a decision to acquit. An erroneous acquittal in either of our cases means that the defendant will be returned to the streets. Again, remember that if the acquittal is erroneous, then the defendant is a repeat offender who has escaped punishment, so unless the defendant's behavioral pattern changes because of either internal or external factors, it is likely that the defendant will offend again. In the case of our accused thief, we might assume that he steals out of economic necessity. Unless his economic circumstances change after his acquittal (a seemingly unlikely event), we have every reason to believe that he will offend again, although he may be more circumspect about his activities. Similarly, our accused rapist may rape because of a psychological condition. In the absence of effective treatment for his condition, it also seems likely that he will offend again.[224]

It is also worth noting that it is possible that there is an additional benefit associated with an accurate acquittal in the hypothetical rape case. Because a conviction for a rape may bring with it the additional possible sanctions of civil commitment or listing under a Megan's law, we might assume that even rape indictments carry with them increased stigma. The fact of an accurate acquittal, in addition to whatever benefits it might give, could decrease the amount of stigma that would attach to

---

[224] There may be additional costs that accrue to the victim in the event of an erroneous acquittal. *See* Shaviro, *supra* note 11, at 532 (citing SUSAN ESTRICH, REAL RAPE 3 (1987)). Of course, it could be that the costs of an erroneous acquittal in the two situations might balance out. This would be so if we assumed that the probabilities of a subsequent apprehension and punishment were precisely calibrated to the amount of future harm. For instance, assume that the amount of harm to society for each theft is c_t, and that the amount of harm to society for each rape is c_r. Also assume that $c_r = 3c_t$. If our rape defendant was three times as likely to be caught after a subsequent crime as our theft defendant, then the total amount of future harm from the two erroneous acquittals would be the same. For this to be true, the likelihood of apprehension for a given crime must be proportionally related to the amount of harm from that crime, but even Gary Becker's deterrence model does not assume that the likelihood of punishment will be set simply in relation to the amount of harm from the offense. *See* Becker, *supra* note 150, at 176-77. At least one other factor to be considered would be the costs of apprehension and conviction. Thus, it is possible that, if it is much more costly to apprehend and convict the perpetrators of rapes than of thefts, then the probability of apprehension and conviction for the two crimes may not be relatively proportional to their social harms. So while it is possible that the costs of erroneous acquittals for the two types of defendants are the same, it is equally, if not more, plausible that they are not the same. The result is that we can at least hypothesize that the utility of an erroneous acquittal will be lower in our hypothetical rape case than it will be in our hypothetical theft case (in other words, the disutility associated with an erroneous acquittal in the rape case is greater than the similar disutility in the theft case).

an innocent defendant from the mere fact of the indictment. This may not be true (or at least as true) of a theft case, meaning that U_{ai} is greater for our rape case than for our theft case.[225]

What is important here is not that each of these differences exists, but that at least some of them do. For so long as there are some real differences between the hypothetical rape and theft cases, then the standard of proof should differ in the two cases. Recall again that in setting the standard of proof, the key is the value of our fraction $(U_{cg} - U_{ag})/(U_{ai} - U_{ci})$. For the standard of proof to be the same in our hypothetical theft and rape cases, we are not required to decide that the values of U_{cg}, U_{ai}, U_{ag}, and U_{ci} are identical in the two cases; instead we just have to show that the values of the four variables in both cases are such that the overall fraction remains constant. I have shown that the utility of accurate verdicts (U_{cg} and U_{ai}) may be higher in our hypothetical rape case than in our theft case, and that the disutility of an erroneous conviction (U_{ci}) is also likely to be lower in our rape case than in our theft case. The disutility of an erroneous conviction (U_{ag}), however, is likely to be greater in our rape case.[226]

Given these relationships, it is at least theoretically possible that the burden of proof values, *SP*(rape) and *SP*(theft) are the same. The likelihood of this, though, seems remote.[227] To do so, we would have to assume that $U_{ai}(\text{rape}) - U_{ci}(\text{rape}) > U_{ai}(\text{theft}) - U_{ci}(\text{theft})$. This would be possible only if we maintain the assumption that $U_{ai}(\text{rape}) > U_{ai}(\text{theft})$. If instead we assume that the utility of an accurate acquittal is the same in the rape and the theft case, then the standard of proof in the two cases will *not* be the same.[228]

Even if we assume that the utility of an accurate acquittal in the rape case is more beneficial than in the theft case, however, the value of the standard of proof for our rape and theft cases will only be the same if the proportional changes in the numerator and denominator of our fractions are the same.[229] Note, though, that the values of the two utility functions

---

[225] It is also possible that there is greater increased utility associated with accurate acquittals in the rape case than in the theft case because an accurate acquittal ensures that the (innocent) defendant will not be erroneously subjected to the additional sanctions of either civil commitment or of the provisions of a Megan's law.

[226] In other words, $U_{cg}(\text{rape}) > U_{cg}(\text{theft})$, $U_{ci}(\text{rape}) > U_{ci}(\text{theft})$, $U_{ai}(\text{rape}) > U_{ai}(\text{theft})$, $U_{ag}(\text{theft}) > U_{ag}(\text{rape})$.

[227] For a proof, please see the Appendix.

[228] For a proof, please see the Appendix.

[229] That is, we would have to show that the proportional increase, n, in $U_{ai}(\text{rape}) - U_{ci}(\text{rape}) = n(U_{ai}(\text{theft}) - U_{ci}(\text{theft}))$ is exactly the same as the proportional increase, m, such that $U_{cg}(\text{rape}) - U_{ag}(\text{rape}) = m(U_{cg}(\text{theft}) - U_{ag}(\text{theft}))$.

in the numerator are moving in *different* directions, while the two utility functions in the denominator are moving in the *same* direction, which would seem to make it very unlikely that the proportional changes will be the same. Instead, it would seem far more likely that the value of the fractions in the rape case will be greater than in the theft case.[230] Remember, because $U_{cg}(\text{rape}) > U_{cg}(\text{theft})$ and $U_{ag}(\text{theft}) > U_{ag}(\text{rape})$, the numerator for the rape hypothetical must be greater than the numerator for the theft hypothetical. All that is then required is that the coefficient of that increase[231] in the numerator of the fraction be greater than the coefficient of increase in the denominator of the fraction, where there need not be any increase at all. While it is possible that the coefficient of increase in the denominator could indeed equal or even exceed that of the numerator, in the absence of more specific information about the exact values of the utility function, it appears more likely that this coefficient will be lower. Accordingly, we should expect that the standard of proof in the hypothetical rape case would be lower than the standard of proof in the hypothetical theft case.[232]

This is not to say that the standard of proof in the hypothetical rape case *should* be lower. My tentative conclusion here is certainly not based on any first order *normative* beliefs about the utility functions for the various verdict possibilities in these two hypothetical cases. Instead, it is based on admittedly speculative assumptions about how we, as a society, might value these various possibilities.[233]

---

[230] This implies that the standard of proof will be lower in the rape case.

[231] By "increase," I am referring to a greater value (if any) in the values of the utility functions for our hypothetical rape case than the values for those same utility functions in our hypothetical theft case.

[232] In other words, because $U_{ci}(\text{rape}) > U_{ci}(\text{theft})$ and $U_{ai}(\text{rape}) > U_{ai}(\text{theft})$, $SP(\text{theft}) > SP(\text{rape})$, because it is likely that $(U_{cg}(\text{rape}) - U_{ag}(\text{rape}))/(U_{ai}(\text{rape}) - U_{ci}(\text{rape})) > (U_{cg}(\text{theft}) - U_{ag}(\text{theft}))/(U_{ai}(\text{theft}) - U_{ci}(\text{theft}))$.

[233] For instance, someone might argue that society really does not associate a greater utility with a rape conviction than a theft conviction (so that $U_{cg}(\text{rape}) = U_{cg}(\text{theft})$), but, because of the existence of civil commitment laws and Megan's laws, there are much greater disutilities associated with erroneous convictions in rape cases and greater utilities associated with accurate acquittals in those cases. Assuming also that $U_{ag}(\text{rape}) = U_{ag}(\text{theft})$, the result is that $(U_{cg}(\text{rape}) - U_{ag}(\text{rape}))/(U_{ai}(\text{rape}) - U_{ci}(\text{rape})) < (U_{cg}(\text{theft}) - U_{ag}(\text{theft}))/(U_{ai}(\text{theft}) - U_{ci}(\text{theft}))$, because $U_{ai}(\text{rape}) - U_{ci}(\text{rape}) > U_{ai}(\text{theft}) - U_{ci}(\text{theft})$. This would mean that the standard of proof in the rape case is higher than in the theft case. I tend to believe that this is wrong, at least as a descriptive matter of society's beliefs, given Nagel et al.'s findings that in rape cases women typically required a lower standard of proof than men, resulting in an overall average that was equivalent to other cases. *See* Nagel et al., *supra* note 56, at 377. In the years since this study, my assumption is that the likely trend has been for society to become even more likely to convict in a rape case than it was then.

Nonetheless, these hypothetical cases demonstrate that in some classes of cases there are sufficient differences in the utilities (and disutilities) associated with various decisions that we would expect society to prefer different standards of proof in those cases. Moving beyond the death penalty example and our hypothetical rape and theft cases, we might suppose that the penalty attached to a crime might often make a difference in the standard of proof society should prefer. For instance, cases involving relatively low-level offenses that almost certainly will not result in any imprisonment perhaps require a lower standard of proof; indeed, they may require a level of certainty that comes close to the preponderance standard. As an example, consider a case involving a traffic offense. Such a case may be technically "criminal,"[234] but the disutilities of an erroneous conviction (because the penalty is so small) will not be ten times the disutility of an erroneous acquittal; instead they are likely to be closer to the 1:1 ratio that is often posited for civil cases.[235] Similarly, the utilities of accurate convictions and acquittals are likely to be close enough together that we can imagine that the resulting value for *SP* should be near (but perhaps a bit above) .50.

Conversely, crimes where the repercussions from the conviction go beyond the formal sentence may be ones in which a higher standard of proof will be required. In these cases, the relative disutility that flows from an erroneous conviction may be sufficiently high so that the resulting standard of proof may be even more than we associate with proof beyond a reasonable doubt. The example I have given of this is capital cases, but it is possible that there are others. Similarly, crimes where the disutility from an erroneous acquittal are relatively high may lead to lower standards of proof, while crimes for which such disutility is relatively low may lead to higher standards of proof.

The examples discussed so far collapse together two reasons why the standard of proof may vary between cases: (a) characteristics about the particular crime and (b) characteristics about the alleged offender. That

---

[234] New York, for instance, defines traffic offenses as offenses (but not crimes) under the penal code, *see* N.Y. PENAL LAW § 10.00 (McKinney 1997), but also requires proof beyond a reasonable doubt for such offenses, *see* N.Y. CRIM. PROC. LAW § 70.20 (McKinney 1992). California is similar: proof beyond a reasonable doubt is required in all "criminal actions," CAL. PENAL CODE § 1096 (West 2002), and criminal actions cover "public offenses," CAL. PENAL CODE § 683 (West 1999), which include infractions, CAL PENAL CODE § 16 (West 1999), and violations of the Vehicle Code are generally infractions, CAL. VEH. CODE § 40000.1 (West 2002). New Jersey is more clear in stating that traffic offenses are not "crimes," *see* N.J. STAT. ANN. § 2C:1-4 (West 1995), but has also held that proof beyond a reasonable doubt applies to such offenses, *see State v. Dively*, 458 A.2d 502, 508 (N.J. 1983).

[235] *See supra* text accompanying note 57.

is, the standard of proof should shift for reasons having to do only with the particular crime that has been charged — either theft or rape — and that if the same defendant was charged with both crimes, the standard of proof applied in the two cases would be different. We have already seen that society may prefer a different standard of proof for different crimes. Just as society places more resources in the hands of law enforcement officials to combat certain crimes, on the assumption that enforcement of those crimes is more important than for other crimes, society may also feel that returning certain types of verdicts is more or less "valuable" in some cases rather than in other cases.

Similarly, the standard of proof may also change simply because of the character of the alleged offender. For instance, in the theft case, the utility of the various possible outcomes may differ depending upon the nature of the person charged with a crime. In the hypothetical, the defendant has engaged in this particular crime several times before. But suppose instead that the defendant was a young person, perhaps in college, for whom this is the first offense. Should we expect the utility values to be the same? The answer is probably no. First, as a matter of substantive law, it is clear that we *punish* repeat offenders more severely than we punish first-time offenders.[236] This suggests, although by no means proves, that society would tend to value accurate convictions more in cases involving repeat offenders than for one-time offenders.[237] Second, as a matter of evidence law, a great deal of information can be admitted at trial so that the juror has the information on which to make distinctions of exactly this sort. Of course, the rules of evidence generally provide that the government cannot provide evidence of the defendant's character to prove guilt.[238] But while evidence of a person's other crimes, wrongs and acts cannot be used to "show action in conformity therewith," it is "admissible for other purposes, such as proof of motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake or accident."[239] The reality is that large amounts of

---

[236] The most obvious example of this is the Federal Sentence Guidelines, which increase the punishment in any given case, based upon the offender's "criminal history."

[237] Of course, this difference in value may simply be accounted for in the amount of punishment, and therefore require no shift in the standard of proof.

[238] *See* FED. R. EVID. 404(a).

[239] *See* FED. R. EVID. 404(b). Furthermore, the defendant himself is allowed to furnish character evidence and, if he does so, the government can then furnish evidence showing the defendant's poor character. *See* FED. R. EVID. 404(a). Finally, evidence of the defendant's past wrongs may be admissible if the defendant testifies at trial. *See* FED. R. EVID. 609(a)(1).

character evidence are in fact admissible at trial.[240] This should alter the level of certainty required (by decision makers) for the standard of proof, based on the character of the defendant.

So, while the rape and theft example could be the product of only one of the two possible distinctions (the specific crime or the specific defendant), I believe that both distinctions are at work. We should expect that the standard of proof will be different between different possible crimes for the same defendant, and that the standard of proof will be different for the same crime for different defendants. Both are sufficient; neither one is necessary.[241]

All this does suggest some additional complications. For instance, it is possible to imagine scenarios in which the utility of an accurate acquittal will be relatively low. Take, for instance, a case in which there is ample evidence that a particular defendant is engaged in the drug trade, or a car theft ring, or stock manipulation, but relatively little evidence that the defendant committed the particular crime with which he is charged. Given the overwhelming evidence that the defendant is a person who is engaged in conduct that is not socially beneficial, we might conclude that even an accurate acquittal, although merited on the evidence, would have disutilities associated with it, because it would free the defendant to continue his wrongful conduct. Allowing the defendant's other wrongful conduct to influence the standard of proof raises concerns with our commitment to concepts such as the exclusion of character evidence,[242] the requirement that the defendant be notified of the charges against him,[243] and other due process issues. Although such possibilities are troubling, the willingness of the law to overlook these concerns is consistent with the rise in the use of certain substantive offenses, such as

---

[240] *See* Allen & Leiter, *supra* note 20, at 1537. In addition to character evidence, the admission of evidence of what Michael Risinger refers to as "heartstrings and gore" (evidence of little evidentiary value except to demonstrate the horrific nature of the particular crime) can only be explained by positing that such evidence is used to make it easier for the prosecution to meet the standard of proof. *See* Risinger, *supra* note 20, at 420-21 n.44 & 444-45. I would rephrase this to say that the evidence comes in and then is used by the decision maker to alter the utilities of the various possible outcomes, such that the value of *SP* is lowered.

[241] There is reason to believe that the character of the defendant, the nature of the crime, and the severity of the penalty all affect verdicts. *See* BENNETT & FELDMAN, *supra* note 20, at 159; Elwork et al., *supra* note 103, at 28.

[242] Despite the reality that much character evidence comes in, our articulated starting point is that character evidence is not admissible. *See generally* FED. R. EVID. 404.

[243] A defendant may not be convicted on the basis of evidence that proves matters that were not alleged in the indictment. *See* United States v. Miller, 471 U.S. 130, 142 (1985).

the federal drug kingpin statute and RICO.[244] For instance, these statutes have themselves greatly increased the amount of evidence that is relevant in a particular case.[245] So, while some aspects of a variable reasonable doubt standard may be worrisome, they are already present throughout our criminal justice system.

Furthermore, it is possible to imagine that in some cases, such a result would even be beneficial. Consider, for instance, a case involving an individual linked to al-Qaeda. The person is arrested and the evidence of a link to the organization is strong, but the evidence of specific criminal wrongdoing is weak. In such a case, there is a risk of an erroneous conviction: the defendant may not have committed the specific crime with which he is charged. Nonetheless, the actual harm that flows from such an erroneous conviction would be low, because of the risk that the defendant would commit other terrorist acts. Accordingly, it would make sense for a juror to require less proof in such a case before convicting.

For now, it is sufficient to note that, given what we know both about decision theory, as well as our society's criminal justice preferences, we can suppose that society will prefer, at least in the abstract, a reasonable doubt standard that is not a fixed point. Instead, society should prefer that a higher standard of proof be used in a case involving the death penalty. We might even think that in some circumstances, for some types of cases, the utility of a decision to convict is sufficiently heightened (or the utility of a decision to acquit is sufficiently lowered) that the preferred standard of proof in those cases should be lower than in others.

B. Applying Variable Reasonable Doubt

The foregoing discussion suggests that, at least in the abstract, we might prefer a reasonable doubt standard that varies depending upon the case. It also assumes that the existing reasonable doubt standard, with its confusing language, is well-designed to allow this to happen. But just because this flexible standard might be preferable in the abstract does not mean that it is preferable in practice.[246] There may be strengths

---

[244] 18 U.S.C. §§ 1961-63 (2000); 21 U.S.C. § 848 (1999).

[245] For a discussion of the increase in the amount of evidence relevant under RICO, see James M. Evans, Note, *"Don't Throw Me Into the Briar Patch": RICO and Rules of Evidence*, 73 NOTRE DAME L. REV. 433, 435, 449-59 (1998).

[246] I am using variable standard here to refer to an instruction that allows the jurors to decide upon the proper standard of proof, either explicitly or because the instruction fails

to a fixed standard of reasonable doubt, requiring a uniform level of certainty across all criminal cases, which would outweigh any advantages that flow from a variable standard. I believe this is not so.

One possible difficulty with a flexible reasonable doubt standard in practice would be the error costs associated with its actual application. As noted in Part II, we can never be sure that real-life jurors apply the standard of proof we would prefer them to use. Assuming for the moment that we want to use different levels of certainty in different cases, how do we know jurors will actually do so? And if the jurors will not, at least in some cases, require the level of certainty that we would prefer, why then should we have a flexible standard?

To start with, once we have concluded that society would prefer, in the abstract, to have a variable standard of reasonable doubt, any system to enforce that standard will create some error costs. There are two types of errors that can arise: the first concerns situations where the jury perfectly applies the burden of proof they are asked to apply, the second concerns situations where the jury fails to properly apply the burden of proof. The question, then, is which approach — a fixed or a variable standard — is likely to minimize the total number of errors, as well as the costs from these errors?

The first type of error is one of over- and under-inclusiveness:[247] where the standard of proof that jurors are asked to apply varies from the

---

to stop the jurors from making this determination. I will discuss the distinction between these two types of instructions, and why we seem to prefer the latter, in Part V. The reference to variable standard, though, does not refer to a situation where a judge might alter the amount of certainty in the instruction based on the type of crime (for instance, requiring .9 certainty in a murder case and .7 certainty in a theft case).

[247] *See* Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 565 & n.13 (1992). Kaplow points out that although it is often assumed that standards do not have this problem, this assumption can be misleading. *Id.* at 565 n.13, 588-89. In Kaplow's terminology, the fixed rule is a simple one: it states that a jury should convict if it has X degree of certainty and that otherwise it should acquit. The variable standard, on the other hand, is complex because it requires the decision maker to make a determination, considering multiple factors, as to what the level of certainty should be. *See id.* at 590-91. The result is that there are over- and under-inclusiveness costs associated with the rule that are not associated with the standard.

One objection to this analysis might be that the variable standard is not a complex standard, but rather a simple standard. For instance, jurors may systematically fail to take into account all the relevant information in setting forth the appropriate level of certainty, resulting in a standard that is chronically simplistic. As a result, the variable standard would suffer from some of the same over- and under-inclusiveness costs of a rule. *Id.* at 594-96. In order to clarify the analysis here, though, I am assuming at this point that the variable standard is perfectly applied. The costs that result from the inexact application of the variable standard will be included in our second group of errors.

standard of proof that we would prefer in a particular case.[248] Suppose, for instance, that we had a reasonable doubt rule that required the jury to have a fixed level of certainty before convicting, and that the amount of certainty required is .90. Maintaining our assumption that we prefer a variable standard of proof (again, depending upon the nature of the case), then the application of this rule will lead to additional erroneous outcomes. Even if we assume that in the vast majority of criminal cases .90 probability is the appropriate value for *SP*, we know that there will be at least some cases in which the preferable standard would either be a higher or lower level of certainty. For example, suppose there is a case where we would prefer a .95 level of certainty. Assuming that jurors actually apply the rigid .90 rule on which the jury is instructed by the judge, there will be a set of cases where we would obtain erroneous convictions: anytime the jurors were between .90 and .95 certain they would end up convicting, even though we would prefer that they would not do so. Of course, the opposite form of error could also occur: there may be cases where erroneous acquittals result because the jury has applied the .90 rule, even though we would have preferred a lower level of certainty, say .80, and the particular case fell somewhere between .80 and .90.

This type of error will not occur with a variable standard. Obviously, jurors are never explicitly informed that the standard is flexible, but because the existing standard is so vague and difficult to understand, as a matter of practice the appropriate level of certainty that jurors will require is left up to the jurors themselves.[249] The fact, though, that jurors are not explicitly told either that the standard is flexible or how to generate the appropriate standard for that particular case, may increase the second type of errors that I will discuss. Because we are assuming that the jurors have properly applied the standard on which they have been instructed, the standard of proof applied should always match the standard of proof society would prefer.

The second type of error occurs when we relax the assumption that jurors perfectly apply the standard of proof on which they are instructed. As I noted above in Part II.B.3, there are good reasons to believe that this assumption does not reflect what jurors actually do. So when jurors misapply the standard of proof, what sort of errors result? Assume for the moment that we have a variable standard of proof. As a result, in a

---

[248] This assumes that jurors perfectly apply the standard of proof on which they are instructed. I will relax that assumption below.

[249] *See supra* notes 130-137 and accompanying text.

particular case the appropriate standard of proof might be .96 (as was the case with the hypothetical standard of proof we calculated for a death penalty case in the preceding section). Under a variable standard of proof, jurors — who have not been explicitly told that they should apply a .96 level of certainty or even told how to come to a decision on the appropriate level of certainty — will be free to set the standard of proof at any level of certainty they prefer. The jurors might, for instance, instead select a level of certainty of .85. As a result, if the level of certainty generated as to guilt during the trial is somewhere between .85 and .96, the jurors will convict, even though, had they properly applied the variable standard of proof, the defendant would have been acquitted.[250] The result is an erroneous conviction. And of course, the opposite might happen as well: if the proper standard of proof in a particular case would be .85, but the jurors instead erroneously apply a .96 standard of proof, then the jurors will erroneously acquit if the certainty generated by the evidence falls anywhere between .85 and .96.

This sort of error can also occur with a fixed standard of proof (although it should happen less frequently as I will discuss below). Return to our previous supposition that the proper standard of proof is .90 and assume that the jury is so instructed. The jury may fail to apply that standard of proof properly and instead may apply a different level of certainty. For instance, in the death penalty example, the jurors may instead apply a lower level of certainty, such as .85. If so, the jurors will then make errors (here, erroneous convictions) in any case where the certainty generated by the evidence falls somewhere between .85 and .90. And again, erroneous acquittals can also result: jurors, having been instructed that the proper standard of proof is .90 may nonetheless apply a higher standard of proof (say .96), which will result in defendants going free even though they should have been convicted under the instruction.

There is reason to believe, though, that these sorts of errors (what I refer to as "misapplication errors") will be less frequent in a legal system with a strictly-defined reasonable doubt rule than in a legal system with a variable standard, if jurors — in reaction to the instruction — vary the

---

[250] Indeed, this may be what happens when death-qualified jurors are used in capital cases. By death-qualifying jurors, the criminal justice system may be selecting jurors who set the level of certainty artificially low. As a result, even though society would prefer a variable standard of proof with a very high level of certainty in such cases, the actual standard of proof used by jurors may be far lower, with the result being erroneous convictions. *See supra* note 214.

standard of proof in a way that is not preferable.[251] This is because it may be simpler for the juror to apply a rule that says "find the defendant guilty if you are ninety percent certain" than to apply a standard that, at best, implicitly invites the juror to decide what the appropriate standard should be. Under the first rule, the juror will only make a mistake if she cannot (or can, but does not) accurately apply the .90 probability rule.[252] Under the variable standard, the juror will also make a mistake if she cannot (or can, but does not) accurately apply the level of certainty required by the variable standard.

The variable standard also creates another possibility of error, what I will refer to as a "miscommunication error": the possibility that the juror will hear the reasonable doubt instruction and then apply a standard that differs from the one we would prefer, because the juror thinks that the standard requires a different level of certainty.[253] This could happen for a number of different reasons. For instance, after hearing the reasonable doubt instruction from the judge, a juror could decide that despite the vagueness in the instruction, she believes that it calls for her to apply a high, fixed level of certainty, in line with the prevailing ideology of reasonable doubt.[254] Perhaps more likely, jurors will recognize, either

---

[251] *See* Kaplow, *supra* note 247, at 586-88. Obviously, the variable standard will not be a pure standard, nor will the fixed rule approach be a pure rule. But the variable approach does appear to give the decision maker more power to determine what the level of certainty should be. *Id.* at 561-62.

[252] Such an error occurs when the juror accurately evaluates the evidence but simply requires either greater or lesser certainty than the rule requires. For instance, the juror may accurately evaluate the evidence as creating .88 certainty, but may make a mistake by requiring only .85 certainty. As noted in Part I.A, errors could arise in other ways, either because of the amount of evidence presented or because the juror misevaluates the evidence presented. *See supra* text accompanying notes 22-29. The effects of such errors from evidence should be the same, regardless of whether a fixed rule or a variable standard is employed. The only exception would be if the variable standard generated a standard of proof that was either generally higher or lower than the fixed rule, in which case there will be differences in the distribution of errors as described in Part I.A. For purposes of this discussion, I assume that a variable standard is not systemically higher or lower than a more fixed rule.

[253] The distinction here is perhaps somewhat artificial, but nonetheless important. Obviously, jurors can misunderstand a standard of proof stated in the form of a fixed rule (such as the preponderance of the evidence rule). In such a situation, the jurors may hear that the rule is a level of certainty of .50, but nonetheless think they are required to apply a higher (or lower) level of certainty. But in the case of a variable reasonable doubt standard, the potential for error is increased, because the jurors may never even hear what level of certainty they should apply. The very vagueness of the standard increases the probability of misapplication error.

[254] Of course, the empirical evidence suggests that jurors generally will *not* do this. *See supra* text accompanying notes 99-102. Nonetheless, the empirical evidence may underestimate the number of people who will apply a high, fixed rule.

consciously or subconsciously, that they are supposed to apply a reasonable doubt standard that varies depending upon the case, but they will fail to come up with the proper standard. The jurors are never given any real instruction in how to come to a decision about the level of certainty, and even if one or some jurors tried to do something like what the decision theory model suggests they might do, there is the possibility that jurors may make mistakes (perhaps even systemic ones) in applying the model (or some variant).[255] Therefore, what we need to worry about with a variable reasonable doubt standard is an increased possibility that the jurors will, without intending to do so, apply either a higher or lower standard of proof than we would prefer.

In choosing between a fixed rule that would apply a consistent standard of proof and a variable standard that would alter the standard of proof, our choice comes down to a trade-off between the costs associated with two types of possible errors. On the one hand, using the fixed standard of proof will create over- and under-inclusiveness errors (that would not exist with a variable standard) where the fixed rule will vary from the level of certainty that we would prefer the jurors to apply. On the other hand, using the variable standard of proof will create errors of misapplication or miscommunication (that would not exist with a fixed rule) where the jurors will fail to apply the proper standard of proof. The question then becomes: which set of error costs will be greater?

The amount of erroneous convictions and acquittals that will flow from over- and under-inclusiveness will turn on the frequency and extent of divergence between the preferred standard of proof and the level of certainty that jurors manifest in their decisions. Assume first that the fixed standard of proof is set at .90. Then assume that the preferred level of certainty in most criminal cases is .90 and the level of certainty never dips above .95 or below .85 even in those few cases where we would prefer the standard of proof to be different. If these assumptions are correct, then there will be relatively few errors of over- or under-inclusiveness, and we might be close to concluding that a fixed rule is preferable to a variable standard. But assume again that the fixed level of certainty is set at .90. Now assume instead that we prefer a .90 level of certainty in only a bare majority of cases, and that the preferred standard of proof can vary anywhere from .75 to .99. If these alternative assumptions are correct, then there may be many errors of under- and

---

[255] *See* Tribe, *Trial by Mathematics*, *supra* note 13, at 1384-86.

over-inclusiveness associated with the fixed rule.[256]

Even if there are a large number of errors associated with over- and under-inclusiveness, we still need to investigate the errors associated with a variable standard. Recall that these errors relate to the greater risk that jurors will misapply a variable standard of proof than a fixed rule. The inquiry as to the magnitude of that risk turns on how much faith we have in the ability of jurors to set accurately the variable standard of proof. For instance, assume again that we would prefer that the level of certainty in a death penalty case be .96. We then need to ask: how often will the jurors, using our present reasonable doubt instructions, actually require that level of certainty? And, when they do not apply that level of certainty, what level of certainty will they apply?

Deciding whether a variable standard as the burden of proof or the fixed rule would be preferable in the end depends on two issues. First, is the preferred standard of proof likely to vary widely in different cases? Second, are jurors likely to apply a standard of proof in a way that varies widely from the one that we would prefer in a given case? If the answer to the first question is yes and the answer to the second question is no, then we should prefer a variable standard of proof. But if the answer to the first question is no and the answer to the second question yes, then we should prefer a fixed rule.

There is no empirical evidence that will directly answer these two questions. Nonetheless, my view is that we should prefer a flexible standard. The analysis in Part IV.A, above, suggests that the preferred standard of proof ought to vary significantly depending on the particular case. Recall that in Part I.C, we saw that different commentators had generated far different standards of proof using the Decision Theory model. For instance, while assuming that the disutility of an erroneous conviction was ten times the disutility of an erroneous acquittal generated a standard of proof that was set at .91, Tribe's ordering of the utilities and disutilities resulted in a required level of certainty of .57 and Milanich's preferred numbers, .55.[257] Meanwhile, my own selected values had resulted in a standard of proof set at .83.[258] This suggests that

---

[256] Indeed, recall the discussion in Part I.A.: the distribution of errors that arises from altering the standard of proof depends critically on a number of factors, including the relative numbers of guilty and innocent defendants who are actually tried, as well as the relative strengths of evidence against those two groups. If we also assume, as some scholars do, that the vast majority of tried cases falls into the narrow band around the standard of proof, then the amount of errors can escalate very quickly. *See* George Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1 (1984).

[257] *See supra* notes 58-66 and accompanying text.

[258] *See supra* text accompanying notes 64-66.

there is likely to be widespread disagreement over the proper values for the utilities, leading to more variation in the standard of proof in particular cases.[259]

Furthermore, other evidence suggests that in actual cases, jurors can apply different standards. In Part II, I discussed a set of studies (the Rank-Order studies) that have attempted to measure, through indirect means, the actual standard applied by subjects. These studies resulted in standards of proof that varied from .525 to .80.[260] While the number of these studies has been limited and the studies themselves have limitations, the variances in their data do suggest that subjects believed that different standards should be used in different cases.[261] Other studies that attempted to measure the standard of proof applied by subjects through different methodologies have also demonstrated a variance in the application of the reasonable doubt standard.[262] Finally, other studies that have asked subjects the level of certainty they would require to convict have also shown variations.[263] These variations in the stated amount of certainty may be the result not just of different sample populations, but also may be the result of the subject's exposure to other stimuli, perhaps within the experiment itself, which have caused them to alter their standard of proof. For instance, some experiments may be conducted against the background of a test case that has altered the subject's perception of the proper level of certainty. Furthermore, subjects in certain parts of the country may have been exposed to certain cases or criminal justice developments that have caused them to alter the utilities of the various verdict choices. To the extent that we can assume that the subjects, at least to a limited extent, are mirroring society's choices, we can then assume that we do prefer a variable standard.

In addition, I would suggest again that the wide variety of conduct made "criminal" by our laws naturally leads to at least the possibility

---

[259] Recall, we are not discussing here what the normative value of reasonable doubt is, which I do not believe we can state. *See supra* Part III. Instead, we are discussing the standard of proof that society descriptively will prefer in particular cases. So the numbers here reflect the wide variance in the preferred standard of proof when subjects are asked directly. *See* Power, *supra* note 9, at 106.

[260] *See supra* text accompanying notes 77-90.

[261] *See also* Stephen E. Fienberg & Mark J. Schervish, *Bayesian Theory and Its Critics: The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decisionmaking*, 66 B.U. L. REV. 771, 779 (1986) (assuming that people accept possibility of variations); ZUCKERMAN, *supra* note 11, at 127 (noting English case holding that reasonable doubt standard is not absolute standard).

[262] *See supra* text accompanying notes 70-76 (describing decision theory calculations in Hastie, *supra* note 22, at 105)

[263] *See* Hastie, *supra* note 22, at 101-02 & table 4.1.

that society should want a varying level of certainty.[264] Recall again our basic equation for the standard of proof:

$$SP = \frac{1}{1 + \left(\frac{(U_{cg} - U_{ag})}{(U_{ai} - U_{ci})} \right)}.$$ [265]

Also recall the discussion of the hypothetical child rape and theft cases. There was reason to believe that we might prefer different standards of proof in those cases because of differences in the utilities (and disutilities) associated with the various possible decisions. The criminal law, though, covers a much wider range of conduct than child rape, theft and murder. It also covers a whole range of minor crimes that differ significantly from these examples. For instance, among the things that various jurisdictions have criminalized are distributing drug paraphernalia,[266] manufacturing burglary tools,[267] fraudulently wearing the Red Cross,[268] and adultery.[269] Each of these examples may seem like a trivial offense that is unlikely ever to be prosecuted, but there is a whole middle range of activity that is criminalized and prosecuted, where the defendant's alleged actions are not as serious as murder or rape, including not only our theft example, but also things like tax fraud, drug distribution, prostitution, and traffic offenses.[270] It seems unlikely that solving Equation 3 will result in the same standard of proof for all of these different types of crimes as for murder. Not only are the punishments vastly different (indicating the likely utilities of accurate convictions are likely to be different), but the disutilities associated with inaccurate convictions and acquittals seem certain to be different as well. While it is possible that these would all balance out in a way that would

---

[264] *See also supra* text accompanying notes 234-235. I am not concluding that a varying standard of proof should be a normative goal in a world of perfect information. But because I do not believe we can obtain perfect information, I believe that it is preferable to vary levels of certainty in different cases.

[265] *See supra* text accompanying note 60.

[266] N.J. STAT. ANN. § 2C:36-3 (West 1995).

[267] N.J. STAT. ANN. § 2C:5-5a & b (West 1995).

[268] 18 U.S.C § 706 (2000). This offense is punishable by imprisonment of up to six months, but under the common understanding of *In re Winship*, it still requires proof beyond a reasonable doubt. *See supra* note 11; Power, *supra* note 9, at 45 n.1.

[269] N.Y. PENAL LAW § 255.17 (McKinney 1999).

[270] Again, although such offenses are often technically not crimes, they are typically covered by the reasonable doubt standard. *See supra* note 234 and accompanying text.

make the value of *SP* constant, it seems to be more likely that there would be some "slippage," such that the value of *SP* will probably be lower in cases involving low-level crimes than in more serious crimes. Given the extraordinary range of different activity that we criminalize, as well as the vast difference in the defendants who come into the criminal justice system, it is quite likely that we prefer a standard of proof that varies significantly from case to case — one that can be either higher or lower than the standard of proof suggested by the traditional view.

Of course, we still need to answer the second question, for even if we want a standard of proof that varies widely, it does not automatically follow that jurors will apply the level of certainty we would prefer. Jurors are, in the end, human beings, and as we have already seen, there are a number of different mistakes that they might make. But for present purposes, the most important inquiry is the risk of miscommunication errors. How likely is it that jurors, after being given a standard reasonable doubt instruction, will then apply a standard of proof in a particular case that varies significantly from the one that we would prefer? Despite our (sometimes) inherent distrust of juries, my belief is that in most criminal cases — *but of course not all* — the jury is likely to do a pretty good job of mirroring the standard of proof that we would prefer.

Note what I do not have to show here: it is not necessary to demonstrate that jurors are likely to reach the normatively correct decision about the level of certainty. Because there is not sufficient information to calculate such a normatively correct level of certainty (either in general or in specific cases), there is no way to ever know if a jury has approximated that result.[271] Instead, what is necessary is just to show that jurors will make decisions about the standard of proof that are similar to those that we would prefer.

Whether this is true depends upon what I mean by "we would prefer." If this is taken to just refer to the level of certainty that some appropriate majority of society had agreed upon, then it is predictable that juries would be good decision makers. The make up of the petit jurors is generally a pretty close approximation of the make up of society.[272] That does not mean that all groups are systemically represented on all juries.[273] Instead, what it means is that the various viewpoints as to the

---

[271] *See supra* Part III.

[272] Again, one obvious exception is death penalty cases. *See supra* note 214.

[273] To the contrary, we do have a tendency to systematically, at least in some cases, exclude certain groups from juries. Kim Taylor-Thompson, *Empty Votes In Jury Deliberation*, 113 HARV. L. REV. 1262, 1276-77 (2000); *see also* Valerie P. Hans & Neil Vidmar, *Jury*

application of the criminal law are generally represented on most juries. This representation of viewpoints may occur even when certain groups are systemically excluded from the jury.[274]

But are the wishes of society an appropriate benchmark for determining the standard of proof in a particular case? An alternative approach to deciding upon the standard of proof might be to suggest this decision should be left to specialists: perhaps judges, or legislators or maybe even academics.[275] Perhaps society should defer to the judgments of these specialists on the grounds that they are better placed to judge among the various costs and benefits in deciding upon a standard of proof. Under this approach, a variable standard of proof would be acceptable only to the extent it reflected the standard of proof preferred by these specialists.

It is by no means clear that the standard of proof preferred by judges would diverge in any significant manner from that standard preferred by society or by jurors. Kalven and Zeisel's work showed that juries are more lenient in criminal cases than judges, which suggests that judges

---

Selection, *in* THE PSYCHOLOGY OF THE COURTROOM, *supra* note 6, at 39, 45 [hereinafter Hans & Vidmar, *Jury Selection*] (noting that present practices do not result in completely representative juries). These effects may be exacerbated by juror decisional rules that do not require unanimity. Taylor-Thompson, *supra*, at 1276-79; *see also* ABRAMSON, *supra* note 126, at 183.

What may be even more telling is that the differences between the make up of society and the make up of juries may also reflect the differences between the make up of society and the make up of voters in our democracy. So while groups are over and under-represented on juries, this may mirror the over- and under-representation of those groups in our other political institutions. If "the outcome that we prefer" is restated as "the outcome that the majority of persons participating in our democracy prefers" (and by using the word "participating," I do not mean to ignore the fact that much "nonparticipation" is involuntary), it may become even clearer why juries are given leeway in deciding the standard of proof: because they are a pretty good representation of the ultimate decision makers in our society.

[274] *See* John M. Conley, William J. Turnier & Mary R. Rose, *The Racial Ecology of the Courtroom: An Experimental Study of Juror Response to the Race of Criminal Defendants*, 2000 WIS. L. REV. 1185, 1186-87 (finding absence of "knee-jerk" racism in jury experiment, but also concluding that outcomes "were responsive to the detailed racial dynamics of individual cases"); *see also* BENNETT & FELDMAN, *supra* note 20, at 166, 169-70 (noting that mere presence of minority defendant may not bias jury); Taylor-Thompson, *supra* note 273, at 1193-95 (summarizing various studies). *But see* Dane & Wrightsman, *supra* note 103, at 106-07 (suggesting that race does matter).

[275] This would echo Laurence Tribe's suggestion that setting the standard of proof is a function that should be left in the hands of the "lawmaker," because the standard of proof should be set not for a particular case, but as a consistent matter for all cases (he also notes that jurors may find it difficult to do this because they do not have sufficient information and, as a result, may not make accurate determinations). Tribe, *Trial by Mathematics*, *supra* note 13, at 1384-86. Allen and Leiter also seem to fall into this camp. *See* Allen & Letier, *supra* note 20, at 1506.

have a lower standard of proof than jurors.[276] Some of this divergence, though, may be the result of the leniency bias that results from juror deliberation; judges do not deliberate with other judges over verdicts, and this might explain their lower standard of proof.[277] Direct surveys of judges' interpretation of what the reasonable doubt standard requires have given results that are generally equivalent to the results of similar surveys of citizens and students.[278] And while there is no data on the other two groups, it is at least imaginable that these groups are no more likely to prefer a standard of proof that is significantly different from what society would prefer.

Furthermore, relying upon the standard of proof that society as a whole would want in a particular case fits neatly with other work on the role of the jury in the American criminal justice system. A group of scholars in recent years has suggested that the strength of our jury systems lies in its ability to give the power of decision making directly to our citizens.[279] These theorists have been focused primarily on the power of juries to acquit over the evidence. But if these theorists are correct that criminal juries serve an important role in giving power directly to citizens to decide upon the proper scope of the substantive criminal law, this presupposes that society as whole, and not judges, are the proper persons to be deciding upon the scope of the substantive criminal law.[280] Similarly, where there is no clear normative basis on which to come to agreement about the standard of proof, why should the standard of proof not reflect the level of certainty society would prefer?

So while juries can make decisions that are not in line with what either the government or perhaps what society writ large might want, they are

---

[276] KALVEN & ZEISEL, *supra* note 7, at 58.

[277] *See, e.g.*, MacCoun & Kerr, *supra* note 72, at 21 (documenting the leniency bias). *But see supra* note 194 for a suggestion that jury deliberation might lower the standard of proof.

[278] *See* Hastie, *supra* note 22, at 102 table 4.1.

[279] *See, e.g.*, ABRAMSON, *supra* note 126, at 2; AKHIL REED AMAR, THE BILL OF RIGHTS 87-88 (1998) (arguing that criminal jury at time of Framers was meant to be populist protection against overreaching governmental officials, and particularly judges); AKHIL REED AMAR, THE CONSTITUTION AND CRIMINAL PROCEDURE: FIRST PRINCIPLES 163 (1997) ("[T]he jury idea was absolutely central to the Founders' Bill of Rights, and their distinctive constitutional idea of popular self-government."); BURNS, *supra* note 20, at 77; HANS & VIDMAR, JUDGING THE JURY, *supra* note 101, at 32; *see also* Albert W. Alschuler, *The Supreme Court and the Jury: Voir Dire, Peremptory Challenges & the Review of Jury Verdicts*, 56 U. CHI. L. REV. 153, 164-65 & n.49 (1989) (noting that juries, although historically less than fully representative of society, have been more representative than other democratic institutions).

[280] *See also Reasonable Doubt*, *supra* note 10, at 1970 (arguing that jurors, as community representatives, are best placed to decide what reasonable doubt means).

likely in most cases to mirror society's preferences, at least at the local level. This in turn suggests that in coming to a decision about the amount of certainty that the jury should require, jurors are likely to apply a standard in line with the one that society would prefer.[281] Thus, absent some influence that would systemically bias the jurors' views of the standard of proof,[282] the result should in most cases be representative of society.

Further abstraction of what constitutes the "preferable" level of certainty may complicate the answer, however. One response to the preceding discussion might be to suggest that the proper level of certainty is not what society would directly agree upon for a particular case; rather the proper standard would be the one which would result from calculating the standard of proof, in the manner of the decision theory model, based on the values that an appropriate majority of society would apply to accurate and erroneous verdicts. In other words, in this version, we would want to discover the values that society would want to place on U_{cg}, U_{ag}, U_{ai}, and U_{ci} and then calculate *SP*.

If this is the definition of "we would prefer," then there may be less reason to believe jurors will apply the correct level of certainty. It is doubtful that jurors actually make calculations about the burden of proof based on some variant of the decision theory model. Indeed, jurors are never given any instructions that would lead them in that direction; and I willingly concede that few, if any jurors, have considered decision theory, let alone the work of John Kaplan and Alan Cullison, either before, during, or after their jury service.[283] So if jurors are doing this, there seems little doubt but that they are not doing it consciously.

If the comparison is between what jurors are likely to do and the standard of proof that would result from the value we (as a society) would provide to each utility function, there may be more reason to believe that there will be some variance as a result of miscommunication errors. Nonetheless, it is possible that, even though jurors are not consciously weighing the utilities of verdict choices, their mental processes are bringing them to the same result *as if* they had done so. The empirical research gives some reason to believe that this is what in fact happens. Some researchers have attempted to rate the "accuracy" of various studies of the reasonable doubt standard by comparing a

---

[281] *See* BURNS, *supra* note 20, at 144 ("In a democratic society, a major discrepancy between legal rules and public mores should not occur too often nor last too long.").

[282] And we have already seen one likely candidate for such an influence in the selection of jurors for death penalty cases.

[283] DeKay, *supra* note 31, at 106-09.

subject's estimated decision criterion with the subject's actual verdict.[284] Interestingly, the research has tended to show that there is a pretty close correlation between the subjects' actual decision on guilt and the decision criterion that would have resulted from the subjects' individual utility values.[285] What is even more interesting is that the decision criterion derived from subjects' utility values was far more "accurate" than the subjects' self-reported decision criterion.[286] In other words, there is good reason to believe that jurors' actual decisions are more consistent with the utility values they would assign in a particular case than with the estimate they would give for the proper standard of proof.[287] Assuming that the utility values of jurors are relatively consistent with those of some appropriate majority of society, then jurors should be relatively accurate in their decisions.

Finally, I believe we may be able to infer something (but perhaps not too much) from over 100 years of consistent practice. From its very inception at the end of the 18th and the beginning of the 19th centuries, the reasonable doubt standard has been vague and thereby invited jurors to apply a different standard in different cases.[288] Despite calls to make the standard more certain, reasonable doubt has remained a poorly-defined term, and one that has relentlessly defied attempts at quantification.[289] Far from being uncomfortable with a shifting and malleable standard of proof in criminal cases, our legal system seems both to welcome the inherent vagueness of the standard and to be hostile to attempts to alter it. Instead, despite some recent calls for clarification of the standard, numerous commentators over the past 100 years have suggested that the standard should remain vague.[290] This consistent

---

[284] For a description of how the process works, *see* MacCoun, *supra* note 68, at 23.

[285] MacCoun reports that Dane found 82% accuracy between the decision-theory derived decision criterion and the actual decision. *See* MacCoun, *supra* note 68, at 27. MacCoun himself found 84% accuracy. *See id.* at 69 table 8.

[286] *See id.* at 23, 69 table 8.

[287] This is also consistent with recent research in the neurosciences.

[288] *See, e.g.*, SHAPIRO, *supra* note 3, at 22-25 (giving examples of 18th century instructions).

[289] *See, e.g.*, Reed v. Roe, No. 95-56020, 1996 U.S. App. LEXIS 27813, at *2 (9th Cir. Oct. 24, 1996) (rejecting use of numerical scale for reasonable doubt); McCullough v. State, 657 P.2d 1157, 1157, 1159 (Nev. 1983) (rejecting lower court's description of reasonable doubt as 7.5 on scale from zero to ten); *see also* United States v. Anglada, 524 F.2d 296, 300 (2d Cir. 1975) (rejecting attempt to characterize reasonable doubt standard as quantitative); State v. Cruz, 639 A.2d 534, 537-38 (Conn. App. Ct. 1994) (describing cases where attempts at quantification had been rejected).

[290] Charles Nesson provides a wealth of such cites. *See* Nesson, *The Evidence or the Event?*, *supra* note 23, at 1365 n.25; Nesson, *Reasonable Doubt and Permissive Inferences*, *supra*

embrace of the vagueness in definition suggests that the reasonable doubt standard, by giving jurors flexibility to demand different levels of certainty in different cases, works in a way that is acceptable to our democracy.[291]

Taken as a whole, the admittedly limited evidence suggests that a variable standard is preferable. We should prefer a variable standard because of the variety of cases and defendants with which the criminal justice system must deal. Furthermore, the errors that result from a variable standard are likely to be lower than the errors that would result from the use of a fixed rule, because the errors of over- and under-inclusiveness should outnumber (and outweigh) the errors of misapplication.

C. Critiquing Variable Reasonable Doubt: The Role of Legitimacy

The analysis thus far has stressed the central role that the costs and benefits that arise from verdicts — both accurate and inaccurate — should play in making decisions about the standard of proof. To some commentators, though, these costs and benefits play (and perhaps ought to play) a much smaller role and that instead the focus should be the extent to which the standard of proof promotes other values in the criminal justice system, the most prominent of which is "legitimacy."[292]

The two main commentators who have suggested the central role of "legitimacy" and related values in explaining the reasonable doubt standard are Laurence Tribe and Charles Nesson.[293] Most importantly

---

note 13, at 1189 n.7 (citing John Henry Wigmore, J.P. McBaine, and drafters of Model Penal Code).

[291] *But see, e.g.*, Shaviro, *supra* note 201, at 552-53 (arguing that assuming that judge-made rules achieve normatively correct outcomes may be misguided because judges and lawyers have incentive for legal system to be perceived positively).

[292] Ronald Allen and Brian Leiter have recently made slightly different and more narrow objections to the use of expected utility theories in setting the standard of proof. First, they raise the problem of the conjunction effect, see Allen & Leiter, *supra* note 20, at 1503-05, a problem that this Article (drawing on others) suggests is overstated, see *supra* note 23. Second, they object that jurors may implement expected utilities that differ from those of policymakers, and presumably therefore the resulting standard of proof will vary. *See* Allen & Leiter, *supra* note 20, at 1505-06. This Article presumes, though, that in a democracy, the important policymakers are "the People" and that the standards of proof generated by jurors are likely to be close to those for society. *See supra* Part IV.B.

[293] *See* Nesson, *The Evidence or the Event?*, *supra* note 23; Nesson, *Reasonable Doubt and Permissive Inferences*, *supra* note 13, at 387-88; Tribe, *An Ounce of Detention*, *supra* note 184, at 1195-97; Tribe, *Trial by Mathematics*, *supra* note 13, at 1330. George Fletcher makes similar claims. *See generally* George P. Fletcher, *Two Kinds of Legal Rules: A Comparative Study of Burden-of-Persuasion Practices in Criminal Cases*, 77 YALE. L.J. 880 (1968). Fletcher appears to believe that society's goal, in deciding whether or not to convict an individual defendant, is

for present purposes, their writings suggest the possibility that, irrespective of any costs and benefits from accurate and inaccurate verdicts, the way in which the current reasonable doubt instruction is implemented can be explained (and perhaps even justified) by the benefits that arise from promoting these other values.[294]

Such claims acknowledge (at least implicitly) that costs and benefits do flow from the inherent vagueness in the reasonable doubt standard, because the present reasonable doubt instruction in practice lessens the burden of proof and therefore increases the number of erroneous convictions.[295] These costs, though, are outweighed by a number of benefits that flow from other values. First, the current reasonable doubt instruction benefits society by expressing society's support for the rights of the individual.[296] Second, the vagueness in the current instruction reinforces the expressive message of the criminal law that committing a crime will lead to punishment; the argument would be that explicit recognition that verdicts just represent probabilities of guilt would result

---

not to protect society, but to reinforce community norms. Furthermore, Fletcher believes that this can only be done in cases where a state can justify convicting by demonstrating the certainty of the defendant's moral blameworthiness. This allows him to conclude that "the interests of the individual ought not readily yield to the supposed benefits of applying the criminal sanctions." *Id.* at 882. For Fletcher, the reasonable doubt standard is grounded in his notion that *only* the morally blameworthy should be punished, without regard to the "positive effects" of convictions. *Id.*

[294] Nesson and Tribe are not completely clear about the extent to which they are making normative claims. Tribe does seem to believe that the current state of affairs is normatively preferable, although he might prefer an instruction that requires even more certainty. *See* Laurence H. Tribe, *A Further Critique of Mathematical Proof*, 84 HARV. L. REV. 1810 (1971) [hereinafter Tribe, *A Further Critique*]. For instance, he is concerned that explicitly recognizing that innocent persons are convicted in the interest of what he refers to as "social gain" would be "wrong," which appears to be a moral claim. Tribe, *An Ounce of Detention*, *supra* note 184, at 386; *see also id.* at 387 (arguing that system that weighed erroneous convictions and acquittals would be intolerable); Tribe, *Trial by Mathematics*, *supra* note 13, at 1381-83 (criticizing decision theory model's failure to include all consequences of verdicts). Nesson is less clear about whether the present system is preferable. His analysis of the functioning of the reasonable doubt standard appears to me (and to him) to be at least partially descriptive. *See* Nesson, *The Evidence or the Event?*, *supra* note 23, at 1391 (apparently disclaiming any normative views). For instance, Nesson states that attempts to either precisely define or to quantify the reasonable doubt standard will undermine that standard's "function" of allowing a social consensus by masking disagreement over the precise quantity of certainty that is necessary to convict someone. *See* Nesson, *Reasonable Doubt and Permissive Inferences*, *supra* note 13, at 1197. Nesson is unclear about whether such masking is to be preferred.

[295] Tribe appears to claim that weighing erroneous convictions and acquittals would be intolerable, but only because of the harm that would flow from this. Tribe does not appear to be making a completely non-consequentialist claim that weighing the costs and benefits would be wrong, even if there were no attendant harms to society.

[296] *See* Tribe, *Trial by Mathematics*, *supra* note 13, at 1374.

in a less effective message.[297] Third, the vague instruction allows for increased acceptance of verdicts because making the reasonable doubt standard more explicit would make it easier for other persons in society to dispute the jury's decision.[298]

All three claimed benefits would suggest that a vague reasonable instruction is preferable because it supports these alternative values. As I discuss in the next Part, I agree that the present instruction is preferable, albeit for different reasons. Where the variable theory and these other commentators part ways is over the value of flexibility. Is it preferable not only that the reasonable doubt instruction is vague, but that the applied standard of proof varies from cases to case? This Article suggests the answer is yes. Focusing on other values, however, might lead to the conclusion that variances in the standard of proof are at best a necessary evil.[299]

None of these claimed benefits of a vague instruction, however, are sufficient to explain the existing regime. First, the idea that the vagueness of the reasonable doubt instruction leads to increased support for the rights of the individual seems highly suspect. The fundamental premise of such a claim is that society would cut back on the protections that individuals obtain in the criminal justice system if it knew that there was not complete certainty that the defendants were guilty. But why should the result not be the exact opposite? Instead, if it was widely acknowledged that defendants are convicted even where there is an absence of complete certainty as to their guilt, the more natural response would be to increase the rights of defendants.[300] The use of other procedural rights for defendants would then lead to a lowering of the number of erroneous convictions. Indeed, the fallout from the revelation of a substantial number of erroneously convicted death row inmates in Illinois led to a complete re-evaluation of the procedures used in death penalty cases there.[301] Similarly, one would imagine that an explicit recognition that a state routinely incarcerates defendants, some of whom

---

[297] *See* Nesson, *The Evidence or the Event?*, *supra* note 23, at 1367.

[298] *See* Nesson, *Reasonable Doubt and Permissive Inferences*, *supra* note 13, at 1195 (arguing that secrecy of jury deliberations and verdict process in general makes it difficult to disagree with verdicts); *see also* Shaviro, *supra* note 201, at 543-44 (classifying claims in similar manner).

[299] *See* Tribe, *A Further Critique*, *supra* note 294, at 1818-19 (noting that amount of certainty should be as high as can be achieved).

[300] *See* Shaviro, *supra* note 201, at 547-48 (noting that explicitness may allow society to openly confront its choices).

[301] Ken Armstrong & Steve Mills, *Ryan Suspends Death Penalty; Illinois First State to Impose Moratorium on Executions*, CHI. TRIB., Jan. 31, 2000, at C1.

are in fact innocent, might lead to additional protections for defendants.[302] Thus, to the extent that the protection of individual rights is a benefit, vagueness in the reasonable doubt standard is an additional *cost* for which we must account.

The second possible benefit — that a vague standard supports the creation of and adherence to behavioral norms — is also problematic. It assumes that verdicts play a role (and perhaps a substantial one) in behavioral norm formation and that this benefit outweighs the attendant costs.[303] But this seems unlikely, because jury trials in the United States criminal justice system are rare events.[304] Over the last few years, approximately eight percent of federal criminal cases have gone to trial.[305] And while there are a few criminal trials that get extensive publicity, the vast majority of criminal trials are resolved with no attention either from the media or the public.[306] Criminal trials have, at times, played a role in behavioral norm formation. The spectacle of the criminal trial in Medieval and Renaissance England may have played exactly this role, because at that time, trials (and the attendant executions) were public displays.[307] But in an era when trials have little public function, it is hard to see how, at least in most cases, they affect norm creation.[308]

---

[302] *See, e.g.*, Givelber, *supra* note 33, at 1395 (arguing that remedying existing inaccuracy in criminal justice system requires more and better investigations).

[303] Nesson is not the first to make this assumption. Justice Brennan, in *In re Winship*, states that the reasonable doubt standard "is indispensable to command the respect and confidence of the community in applications of the criminal law [because] [i]t is critical that the moral force of the criminal law not be diluted by a standard of proof that leaves people in doubt whether innocent men are being condemned." 397 U.S. 358, 364 (1970). The criticisms of Nesson's position apply with equal force to Justice Brennan's. Ronald Allen states, though, that this rationale did not survive the Court's subsequent decisions in *Mullaney v. Wilbur* and *Patterson v. New York*. *See* Allen, *Restoration*, *supra* note 11, at 41.

[304] *See also supra* notes 165-166 and accompanying text.

[305] In Fiscal Years 1996 to 1998, between 6.9% and 8.2% of federal criminal defendants were either acquitted or convicted at trial. *See* BUREAU OF JUSTICE STATISTICS 1998, *supra* note 35, at 410 table 5.23; BUREAU OF JUSTICE STATISTICS 1997, *supra* note 35, at 400, table 5.23; BUREAU OF JUSTICE STATISTICS 1996, *supra* note 35, at 450, table 5.29.

[306] *See also* Allen, *Reconceptualization*, *supra* note 20, at 431 n.69 (making similar point).

[307] *See* P.G. Lawson, *Lawless Juries? The Composition and Behavior of Hertfordshire Juries, 1573-1624*, *in* TWELVE GOOD MEN AND TRUE, *supra* note 206, at 117, 148. Paul Mahoney has suggested that by the 18th Century, public executions no longer played this function and indeed actually *undermined* the formation of the very behavioral norms that the English government desired. *See* Paul G. Mahoney, *Norms and Signals: Some Skeptical Observations*, 36 U. RICH. L. REV. 387, 399-402 (2002).

[308] *But see* Kahan, *Secret Ambition*, *supra* note 160, at 421-22 (giving examples of potential norm formation (or perhaps norm reinforcement) by public statements associated with particular trials); McAdams, *An Attitudinal Theory*, *supra* note 160, at 374-78.

This is not to say that the criminal law does not operate, at least in part, through expressive functions, nor am I denying that criminal trials can play a role in behavioral norm formation. As discussed in Part III, several commentators have lately been discussing the expressive role of the criminal law. These commentators have generally not pointed to *trials* as the means by which society expresses its disapproval of various actions. Recall, for instance, the social norms theory that law can alter people's behavior by changing their beliefs about which behaviors will generate approval from society.[309] Richard McAdams' explanation of how a governmental entity will accomplish this depends upon the public having some knowledge of the governmental entity's actions. Without such knowledge, there is no signal for people to internalize. Indeed, when McAdams describes the actions of legislatures, he notes *"a law the public generally fails to notice can hardly be expected to have an expressive effect."*[310] Criminal trials, because they generally go unnoticed by society, will have little expressive effect.[311] Therefore, the legitimacy benefits gained by a vague instruction seem quite limited and cannot serve as an adequate normative explanation of the standard.

Even though the actual litigated cases in which the law is "expressed" to society as a whole are quite small, this claim might hold up if we could show that the benefits that society accrues in these few cases where there is effective expression, outweigh the costs of the vague

---

[309] McAdams, *An Attitudinal Theory*, *supra* note 160, at 343-49. In a related article, McAdams has suggested that the law can solve coordination problems for parties. *See* Richard H. McAdams, *A Focal Point Theory of Expressive Law*, 86 VA. L. REV. 1649 (2000). Like the Attitudinal Theory, the Focal Point Theory depends on the public knowledge of the expression by the "law," something that this Article contends is missing from most jury verdicts. *See e.g.*, *id.* at 1667-68.

[310] McAdams, *An Attitudinal Theory*, *supra* note 160, at 362.

[311] Of course, some criminal juries may have an effect on the way the public views the law. Dan Kahan's theory of expressive criminal law also leaves little room for the average criminal jury; instead the main player is the legislature and the executive. *See* Kahan, *Secret Ambition*, *supra* note 160, at 442-45 (describing executive branch and legislative branch actions as expressing society's commitment to death penalty; sole exception listed is election in which members of California's Supreme Court were not retained by public); *id.* at 460-62 (describing attempt of gun control proponents and opponents through use of legislative enactments); *id.* at 463-67 (noting use of expressive arguments to support and oppose hate crime laws). But particular cases can have expressive effects. Kahan points to an example of a particular case from Texas where a judge granted a lenient sentence to the convicted killer of two homosexuals as a way of expressing his "low valuation of homosexuals," *id.* at 467, a decision which was then attacked by others as "send[ing] messages to the community that it's still open season on gay and lesbian citizens." *Id.* (quoting Lisa Belkin, *Texas Judge Eases Sentence for Killer of a Homosexual*, N.Y. TIMES, Dec. 17, 1988, at 8). Furthermore, any particular case can have an expressive effect on the small group that is aware of the decision. *See supra* note 166.

reasonable doubt instruction. Assuming, as at least Tribe does, that a high level of certainty is preferable, a vague reasonable doubt instruction results in substantial costs.[312] To the extent that a vague reasonable doubt instruction allows jurors to apply a decision criterion that is below that which the legitimacy theorists would otherwise prefer, the vague reasonable doubt instruction results in a number of erroneous convictions.[313] So this claim depends on a doubtful proposition: that the costs from these erroneous convictions are outweighed by the benefits that accrue to society from "expressing" the law in the few cases of which society is aware.[314]

This leaves us with the third alleged benefit: that the vague reasonable doubt instruction promotes social welfare because it increases the acceptance of verdicts. This claim presupposes that a more explicit reasonable doubt rule (either one that set forth an explicit decision criterion or one that explicitly stated that jurors should come to their own decision about the proper criterion) would make it easier for other persons in society to dispute the jury's decision. As a result, social acceptance of the verdicts would be lessened. Like the previous claim, though, this claim requires general knowledge of the results of criminal trials.

There are, of course, exceptions to this generalization. The public is aware of some criminal trials, and in those cases it is possible to imagine that a vague reasonable doubt instruction has the benefit of increasing social acceptance of verdicts. Because the public is aware of few of these

---

[312] *See supra* note 294 (discussing the extent to which Tribe contends that a high level of certainty is normatively preferable); *see also* Fletcher, *supra* note 293, at 882 (also claiming that high level of certainty is preferable). Again, Nesson's claims about the reasonable doubt standard appear to be descriptive, not normative, so it is not clear whether Nesson would agree with Tribe that the reasonable doubt standard, in the absence of the other benefits he claims, should (or should not) require a high level of certainty.

[313] Recall that the empirical evidence strongly supports the conclusion that jurors, at least sometimes, require far less than .95 or even .90 certainty before convicting. *See supra* notes 70-103 and accompanying text.

[314] A weaker version of this claim could be made: the expressive function of the law in a particular case is not to "express" the law to society at large, but rather to the participants in the particular case — the defendant, the lawyers, the witness, the judge, and the jurors. If this is the claim, though, it does nothing more to justify a vague reasonable doubt instruction than the stronger claim. Participants in the case, unlike society at large, will actually see the reasonable doubt standard in practice. And what they will experience is a standard that claims to require a very high level of certainty, but that in actuality can (in some circumstances) require much less. The result will not be to reinforce the expressive message of the criminal law, but to undermine it. When participants come to realize that defendants are punished based on less than complete certainty, they may come to view the criminal law as simply a wager on the possibility of the defendant's guilt.

cases, to prove this claim the legitimacy theorists are forced to assume that the benefits of creating public acceptance of the guilty verdicts in the few cases of which there is public awareness outweigh the error costs that arise from having a vague reasonable doubt standard. If the public is unaware of the results of most criminal trials, there is nothing for the public to accept or reject. Again, the benefits seem too small to justify the standard.

A variant on this acceptance thesis might be more appealing. Professor Meir Dan-Cohen has suggested that there are two types of legal rules.[315] The first are "conduct rules," which are addressed to the general public and are designed to guide the conduct of members of the public; the second are "decision rules," which are addressed to officials and are designed to guide their decisions about the conduct of the general public.[316] Professor Dan-Cohen then asks us to "[i]magine . . . that each of the two groups occupies a different, acoustically sealed chamber. This condition I shall call 'acoustic separation.'"[317] According to Dan-Cohen, in such a system the conduct rule and the decision rule could be different. For instance, we could have a conduct rule that tells the general public that duress is never a defense to a crime (thus, no one will ever rely on it in making a decision), but we could also have a decision rule that allows the defense of duress.[318] Of course, in the real world, true acoustic separation does not exist. Nonetheless, Dan-Cohen argues that some acoustic separation does exist, such that there is selective transmission of certain legal rules to various groups and that this selective transmission explains various doctrines in criminal law, such as duress, the duty to testify, and ignorance of the law, to name a few.

It is possible to imagine that partial acoustic separation exists in connection with reasonable doubt. The general public receives a "conduct rule" that states that no one will be punished unless the jury is very certain that the person is guilty. The officials, though, receive a different "decision rule": the actual instruction (implicitly) leaves it up to the jurors to decide what is the proper level of certainty for a

---

[315] Meir Dan-Cohen, *Decision Rules and Conduct Rules: On Acoustic Separation in Criminal Law*, 97 HARV. L. REV. 625 (1984). For a critique of Dan-Cohen's article, see Richard Singer, *On Classism and Dissonance in the Criminal Law: A Reply to Professor Meir Dan-Cohen*, 77 J. CRIM. L. & CRIMINOLOGY 69 (1986).

[316] *See* Dan-Cohen, *supra* note 315, at 630.

[317] *Id.*

[318] *See id.* at 633.

conviction.[319] On this account, legitimacy is created for the criminal justice system as a whole by making it appear that only the guilty are imprisoned, while in reality the other goals of the system are achieved by requiring less certainty in particular cases.

This account has some difficulties. True acoustic separation cannot exist: there is bound to be some leakage whereby members of the public become aware that the applied standard of proof is lower (at least occasionally). So for those people, there will be no gain from selective transmission. Furthermore, selective transmission of this sort has potentially both positive and negative effects. On the one hand, the general law abiding public receives an assurance from the conduct rule that it is unlikely to be wrongly convicted; on the other hand, criminals receive the message that they are unlikely to be convicted.[320] Are we better off reassuring the public that it will not be wrongly convicted or should we instead inform criminals that the absence of complete certainty will not stop us from convicting them? It is by no means clear that selective transmission is likely to be an independent normative justification for our current reasonable doubt regime. Nonetheless, I do believe that it helps explain in part our current reasonable doubt instruction, as I will explain in Part V.

In short, other values, including legitimacy, do not provide sufficient normative justification for the existing reasonable doubt standard. While there are some benefits that arise from these other values, these arise only in the small number of cases of which there is public awareness. Much more widespread are the costs to which these values would also point: costs from defendants erroneously convicted when the jurors apply a lower standard, costs of inaccurate verdicts that arise from too narrow conceptions of defendants' rights, and finally costs that result from foregoing whatever additional benefits accrue from giving rights. Other values, as presently stated, cannot justify a vague reasonable doubt instruction.

---

[319] Note that here, conduct rules and decision rules are in a sense reversed from their alignment in Dan-Cohen's article. There conduct rules were meant to make the criminal law appear broad, and thereby restrict the behavior of the general public, while decision rules tended to narrow the scope of the criminal law to avoid bad results in particular cases. In my example, the reasonable doubt conduct rule makes a conviction appear harder to obtain than the decision rule actually makes it.

[320] *See* Singer, *supra* note 315, at 86-87 (noting that acoustic separation would result in different messages being given to general public and to professional criminals).

V. THE VAGUENESS DILEMMA: JURY INSTRUCTIONS

The question that remains is: why have a vague instruction? The discussion to this point has suggested that there are good reasons to want a reasonable doubt standard that is flexible in setting the standard of proof for a particular case. The evidence suggests that we should prefer a variable standard in the abstract because criminal cases do differ. In practice, we should prefer the variable standard because jurors are more likely to reach preferable outcomes under a variable standard than they are under a fixed rule.

None of this answers, however, why the reasonable doubt instruction itself is so vague and unintelligible. The existing reasonable doubt instruction, by either saying nothing about what reasonable doubt means or by being unintelligible in its definition of reasonable doubt, permits the instruction to be flexible, but by no means mandates such flexibility. If what we want is a variable standard, where jurors adjust the standard of proof depending on the facts of each case or crime, then why not just tell jurors that? Why do we give jurors an instruction that *allows* them to vary the standard (by telling them close to nothing), rather than *explicitly instructing* them to vary the standard of proof?

Furthermore, the previous part suggests that errors will persist so long as we have a vague reasonable doubt instruction. There will be inaccurate convictions because jurors, after hearing the vague instruction (which does not explicitly tell them to vary the standard of proof), will misapply the instruction and require less certainty than society would prefer; conversely, inaccurate acquittals will also now occur because jurors will in some cases require a greater level of certainty than society would prefer.[321] The result is that there will still be errors that could have

---

[321] The total amount of error costs is likely to be less than in a condition where we assume that the preferred level of certainty is very high. To the extent that we assume (admittedly, somewhat artificially) that the switch from preferring a high level of certainty to a variable level of certainty will result in a 1:1 tradeoff of erroneous convictions for erroneous acquittals, almost everyone would agree that the net social costs would decrease. (Because just about everyone assumes that erroneous convictions are more costly than erroneous acquittals. *See supra* notes 46-58, and accompanying text). In addition, it also seems likely that the total errors will decrease. This is because the result of a vague instruction (obviously, in conjunction with the absence of other constraints and jurors' maximization of their own interest, as discussed in Part III, *supra*) is that jurors apply a level of certainty somewhere below .90, say on average .80. Acknowledging that the proper level of certainty is also below .90, say on average .85, should result in a decrease of the overall number of errors as there will be more cases where the jurors correctly apply the reasonable doubt standard. (If, however, jurors are systematically misapplying the level of certainty, then this assumption might not be true. But, as I claimed in Part IV.B, *supra*, I believe that jurors are likely, in most cases, to approximate relatively precisely the

been avoided by being explicit, rather than vague, about what the reasonable doubt standard requires.

Given the likelihood of the costs of a vague instruction and the large number of trials in which such errors can occur, there does not appear to be any reason to prefer a vague instruction over one that explicitly invites the jurors to vary the standard of proof. So why does the vague reasonable doubt instruction persist?

The main benefit of a vague reasonable doubt instruction is that it avoids a possible quagmire for the criminal justice system: just how would a judge explicitly instruct jurors to calculate their own standard of proof? Such an instruction would have to begin by telling the jurors that the reasonable doubt standard is not a fixed line, but rather a level of certainty that is different from case to case. The instruction would then have to tell the jurors what factors to consider in setting the level of certainty: the benefits of accurate verdicts and the costs of inaccurate verdicts. Of course, the instruction would not be able to tell the jurors precisely what these costs and benefits are — for if this was possible, then there would seem to be a way to calculate *SP* directly and there would be no need for a variable standard. Instead, the instruction might look something like the following:

> What is a reasonable doubt? It represents the degree of certainty that you believe should be required of the government in this case, after considering the possible benefits of an accurate determination and the costs of an inaccurate determination in this particular case. In other words, you should consider both the possibility that your verdict will be correct and that your verdict will be wrong, and the resulting costs and benefits from such possibilities. After weighing those possibilities, you should come to your own decision about what degree of certainty you must have before voting to convict the defendant. If, after you decide on this level of certainty, you do not believe there is a higher degree of certainty that the defendant is guilty, you must acquit, for the government has not met its burden of persuasion. On the other hand, if, after deciding on the appropriate level of certainty, you do believe there is a higher degree of certainty that the defendant is guilty, you must convict.[322]

As I believe this example demonstrates, drafting language concisely communicating to the jury the idea behind the theory of variable reasonable doubt is quite difficult. Instead, any meaningful instruction

---

level of certainty that society would prefer.)

[322] For another variant on this, see Nagel et al., *supra* note 56, at 380.

into what the jury should do to determine the standard of proof seems likely to fall prey to the problems that bedevil jury instructions in general: low levels of clarity and understanding.[323] The resulting instruction, if no more understandable than the current instructions, might have no added benefit over not instructing at all (or continuing to use the present incomprehensible instructions). In other words, there may likely be few, if any benefits from attempts to be more specific in explicitly telling jurors to vary the standard of proof.

Furthermore, even if language could be crafted that would make the instruction comprehensible to jurors, it seems unlikely that the instruction would stay that way for long. As I noted, normative agreement about the bases of criminal law is elusive.[324] For that reason, the instruction could never tell jurors precisely which utilities (and disutilities) to weigh — again, agreement on this would lead directly to a precise value for the standard of proof. Instead, the instruction would have to remain ambiguous about what these utilities are. Presumably, though, some interest groups in the criminal justice system, particularly prosecutors and defense attorneys, would have good reason to attempt to create greater specification of these costs and benefits. The natural result would be an endless series of arguments about which costs (erroneous convictions? erroneous acquittals? dangers of serial offenders? effects of imprisonment on innocent individuals?) and which benefits (deterrence? rehabilitation? incapacitation?) should count and be explained to the jury. The vague instruction avoids these conflicts by simply ignoring them.

A similar concern also explains the continuing judicial hostility to attempts to quantify the reasonable doubt standard.[325] Explicit discussion of the amount of certainty, although it would lead to a more predictable standard, opens up questions about *what* the level of certainty should be. As we have seen, societal consensus on this issue is very unlikely — the commentators cannot even come to agreement about what they believe the proper ratio of erroneous convictions and acquittals should be.[326] By refusing to allow any quantification of the reasonable doubt standard, the judges stifle debate on an issue that can never be satisfactorily resolved.[327] It is not that legitimacy of verdicts is

---

[323] *See supra* text accompanying notes 126-137.

[324] *See supra* text accompanying notes 146-200.

[325] *See supra* note 289 and accompanying text.

[326] *See supra* text accompanying notes 46-47.

[327] Daryl Brown's recent suggestion that courts adopt a rule allowing defendants to choose which reasonable doubt instruction might illustrate this point. *See* Brown, *supra*

increased as a result of not quantifying the standard; as I noted in Part IV.C, it is quite debatable whether vagueness increases or decreases legitimacy.[328] Instead, there is simply an avoidance of social losses connected to pointless debates trying to achieve a stable decision about the level of certainty necessary when no such stable state is possible.[329]

A way to avoid these problems might be to limit the instruction to a simple invitation to the jury to vary the burden, with no direction as to how to accomplish this. For instance, the instruction might be limited to the following:

> The government, represented by the prosecutor in this case, has the burden of establishing and proving the guilt of the defendant beyond a reasonable doubt. If, after you have heard all of the evidence and applied to it the rules of law on which I have instructed you, you are not convinced of the defendant's guilt beyond a reasonable doubt, you must acquit the defendant. If, on the other hand, you are convinced of the defendant's guilt beyond a reasonable doubt, you must convict the defendant. The question, naturally, is what is a reasonable doubt? It represents the degree of certainty that you believe should be required of the government in this case.

This instruction would have the benefit of being easier to comprehend and would directly communicate the idea of varying the level of certainty without telling the jurors *how* to do it. The failure to direct the jurors in the means by which to arrive at the level of certainty would seem relatively costless: presumably, the absence of such direction would result in no more errors than an instruction that does give such direction (but as a result becomes incomprehensible) and would have the benefit of reducing the opportunities for prosecutors and defense attorneys to compete over which costs and benefits should count. So why shouldn't such an instruction be preferable?

---

note 130, at 1129. Granting control on this issue to defendants will create an incentive for defense counsel to attempt to alter the instruction to emphasize how difficult the standard is to meet. Although I question whether such alterations would really have the desired impact, *see infra* text accompanying note 350, one likely consequence is that prosecutors, when discussing how they have lost cases, will discuss how unreasonably high the standard is. The result of such discussions might well be a public backlash against the underlying value choice. This might lead either to a return to the vague instruction or to a more specific prosecutor-friendly instruction.

[328] *See supra* text accompanying notes 292-320.

[329] Here I am drawing on Kahan's point about deterrence: that talk about deterrence masks social disagreements about the expressive nature of criminal law. *See* Kahan, *Secret Ambition*, *supra* note 160, at 476-77.

The difficulty is that such an instruction, which invites flexibility but gives no instruction in how to vary the level of certainty, would be unstable. Once judges tell jurors that they should vary the level of certainty, prosecutors and defense attorneys would find it in their interests to make closing arguments that jurors should vary it in the direction that favors the government or the defendant for any number of reasons. Such arguments would naturally lead to attempts by both prosecutors and defense attorneys to have the judge instruct defendants about their costs and benefits.[330]

Nonetheless, if such an instruction could be stable, this instruction would seem to be preferable to the present vague instruction, if only because the instruction would ensure that jurors would actually do what seems socially preferable: vary the standard of proof. There are at least two reasons why a vague instruction might still be preferable.

First, vagueness may allow reasonable doubt to have different meanings in different communities; this is a variant on the "acoustic separation" argument discussed in Part IV.C. By being unclear about the precise amount of certainty that is required for a conviction, the instruction may allow the general public to believe (incorrectly) that it runs little risk of being wrongly convicted if ever haled into court (the conduct rule). Professional criminals, who have experienced the criminal justice system up close, will be transmitted a different message: that proof beyond a reasonable doubt can result in convictions even when the evidence is not overwhelming (the decision rule). In theory, this is a good thing, because it helps to deter crime while leaving most people feeling secure, but in practice it can have unsettling implications. There is certain to be some leakage from professional (or even unprofessional) criminals to their families and communities of the variable nature of reasonable doubt. And because defendants in the criminal justice system are disproportionately of lower economic status and from minority communities, this can lead to people in those communities feeling as if they are subject to a different form of justice than the general population as a whole. Nonetheless the existence of the vague instruction, by promoting the conduct rule, may help to stem some of the social losses that would come from explicit acknowledgment of the actual decision rule.

---

[330] A contemporary example of this is the attempt to get explicit instructions about which things matter in death penalty sentencing. *See, e.g.*, 5 WAYNE R. LAFAVE ET AL., CRIMINAL PROCEDURE 700-01 & n.21 (2d ed. 1999) (noting restrictions on jury instructions at sentencing phase).

Second, up to now, I have assumed that the level of certainty applied by jurors is likely to be the same, or close to the same, that society would prefer.[331] Perhaps a more realistic assumption is that jurors will apply a standard of proof that is close to, but generally lower, than the burden that society would prefer. The conclusion that jurors, left simply to their own devices, might set the standard of proof too low, is supported by several observations. In Part II, we saw that Kahneman and Tversky's prospect theory suggests that jurors' calculation of the standard of proof might be systematically low because of the way in which jurors' would calculate the standard of proof.[332] In particular, jurors might tend to overvalue convictions and undervalue acquittals. What distinguishes actual jurors from society as a whole is that they choose to follow the law and serve, rather than take advantage of multiple opportunities not to serve. Actual jurors, therefore, may have a greater preference for obeying the law and have less familiarity with those who do not. As a result, they may overvalue convictions and undervalue acquittals. This effect may be enough to decrease the value that actual jurors place on *SP*. If jurors are in fact systemically (slightly) undervaluing *SP*, then the present instruction may help correct for this by placing a slight upward pressure on jurors' evaluation of *SP* by not telling jurors the level of certainty that they should apply. In other words, while the instruction may have little ability to place formal or informal constraints on jurors' behavior, the instructions' vague command that jurors apply a high level of certainty may serve, particularly in conjunction with defense lawyers' arguments, to raise jurors' threshold of conviction. And this may be particularly true with the sort of law-abiding people who actually turn up for jury service. In addition, this rhetorical function of the instruction may help explain why deliberating jurors generally raise the standard of proof.[333]

Both of these observations suggest that an additional benefit to the vague instruction might be its ability to increase the value of *SP* compared to the simple, clear instruction. The vague reasonable doubt instruction, while it by no means mandates a high level of certainty, may have the ability, compared to an instruction that explicitly asks the jurors to vary the standard of proof, to encourage jurors to apply a slightly higher value of *SP*. All of this is somewhat speculative. It appears,

---

[331] *See supra* text accompanying notes 272-274.

[332] As we saw in Part II, jurors, in weighing their own prospects, should systemically undervalue *SP*.

[333] *See* MacCoun & Kerr, *supra* note 72, at 30.

though, that there are few benefits to be gained from attempting greater specificity in the instruction, and good reason to suspect that there would be larger offsetting costs.

The next question then is whether the current instruction should be changed, as some have argued.[334] The existing instructions protect the ability of jurors to vary the standard of proof without exposing the instruction to actual manipulation by the various other players in the criminal justice system. Any attempt to be more specific about jurors' task would be problematic, because it would lead prosecutors, defense lawyers, and judges to attempt to manipulate the instruction to recognize some, but not other, "important" utilities.[335] This suggests that improvements, if any, should not try to make the instruction more specific.

Nonetheless, perhaps steps should be taken to improve the ability of the instruction to move jurors' level of certainty upward. This might be desirable if, despite the upward pressure present in current instructions, jurors are still systemically undervaluing *SP* and the instructions could be altered to place more pressure on jurors to raise the value of *SP*. I am skeptical that (outside the context of the death penalty) such systemic undervaluing of *SP* presently occurs. I believe that the decisions of jurors, in most cases, are likely to be a pretty good surrogate for the decisions that society as a whole would make.[336] Furthermore, the near universal present acceptance of the present legal regime — which has existed now for over a century — suggests to me that it is close to what we would prefer. This stability is remarkable when one compares it with other aspects of criminal procedure, which have seen some remarkable changes over the past forty years, let alone the past 100 years. Indeed, the acceptance of proof beyond a reasonable doubt is all the more exceptional when one considers that it is not even mentioned in the Constitution. Despite its omission from that document, not even the most textualist of the present justices on the Supreme Court has suggested that the need for proof beyond a reasonable doubt is not constitutionally required; instead, one of these justices has been among the most protective of the right.[337] The complete lack of controversy that

---

[334] *See supra* note 9.

[335] *See supra* text accompanying notes 324-329.

[336] *See supra* text accompanying notes 272-274.

[337] Justice Scalia has been the justice who has been the most protective of the defendant's right to a jury decision as to each element of the offense. *See* Neder v. United States, 527 U.S. 1, 30-40 (1999) (Scalia, J, dissenting); California v. Roy, 519 U.S. 2, 6-8 (1996) (Scalia, J., concurring).

surrounds the reasonable doubt standard implies that we are pretty happy with the way the standard operates. Of course, juries may be systemically undervaluing *SP* compared to some separate normative ideal, but because I do not believe such an independent normative ideal exists, I do not see this as a problem.

Nonetheless, there might be some reason to believe that juror undervaluing of *SP* has gone relatively unnoticed (and thus has escaped controversy) and that jurors are not quite the ideal I have suggested. If so, the question then becomes whether changes to the reasonable doubt instruction would be effective. As I noted in Part II, jurors routinely fail to understand the instructions — because the instructions are confusing and often only given once, in oral form — and, even when they understand the instructions, jurors may well ignore the instructions if they suggest the juror do something that is not in his interest.[338] Revised reasonable doubt instructions will have to show that they are both more informative and more likely to be retained by the jurors.[339] The approach most frequently considered would be to maintain the structure of the present instruction, but clarify it. This approach might result in the adoption of something like the revised Federal Judicial Center instruction, which defines reasonable doubt as "proof that leaves you firmly convinced of the defendant's guilt."[340] Lawrence Solan has suggested that this instruction would be easier for jurors to understand and apply.[341]

Even if jurors comprehend this revised instruction (or others), they may not apply it. I have suggested that jurors have strong reasons to apply a standard that *they* prefer, rather than the standard that *we* prefer. Changes to the instruction will have to overcome the jurors own self-interest in applying the standard that they would prefer. The empirical evidence suggests that it is possible, but difficult, to overcome jurors' own interests.

Studies on the effects of the Federal Judicial Center instructions are a mixed bag. Irwin Horowitz and Laird Kirkpatrick tested several different jury instructions with two different versions of a mock case,

---

[338] *See supra* notes 126-129 and accompanying text.

[339] One reform that should have some effect would be more readily providing written copies of the instructions. Note, though, that in many experimental settings the instructions are written and are still not effective constraints on subjects' behavior.

[340] Federal Judicial Center, PATTERN CRIMINAL JURY INSTRUCTIONS No. 21, at 28 (1988).

[341] *See* Solan, *supra* note 9, at 144-46; *see also* Brown, *supra* note 130, at 1107 (arguing for more clarity).

using jury-eligible adults as their subjects.[342] They found that the Federal Judicial Center instruction led to statistically significantly fewer convictions for the "weaker" of the two mock cases but no effect in the "stronger" case.[343] A more recent study by Chantal Mess Koch and Dennis Devine tested the Federal Judicial Center instruction against an instruction that did not define reasonable doubt and found that neither instruction made a significant difference on jury decision making.[344] Despite the apparent inconsistency between this study and that of Horowitz and Kirkpatrick, the difference might be ascribable to the strength of the evidence Koch and Devine presented to their subjects, which they describe as "moderate."[345] Of course, the differences could be the result of other factors as well. Nonetheless, the preliminary data gives some reason to believe that changes to jury instructions will have little effect in most cases, but might have some effect in a small number of cases where the evidence is close.

Real world data also gives little support for the benefit of changing the reasonable doubt instruction. Lawrence Solan collected data following New Jersey's adoption of a reasonable doubt instruction similar to the Federal Judicial Center instruction in *State v. Medina*.[346] Solan admits that

---

[342] Irwin A. Horowitz & Laird C. Kirkpatrick, *A Concept in Search of a Definition: The Effects of Reasonable Doubt Instructions on Certainty of Guilt Standards and Jury Verdicts*, 20 LAW & HUM. BEHAV. 655 (1996).

[343] *Id.* at 662-63 & table I. Indeed, the Federal Judicial Center instruction led to a statistically significant higher mean "guilt score" in the stronger case than the two existing instructions tested. "Guilt scores" were the subject's rating, on a 1-6 scale, of the defendant's likelihood of guilt, where 1 indicated the subject's certainty of not guilty verdict and 6 represents certainty in a guilty verdict. The higher the score, the more certain that the defendant is guilty. *Id.* at 662. In addition, the Federal Judicial Center instruction also resulted in more guilty verdicts than two of the other instructions. *Id.* at 663 table I.

[344] Chantal Mees Koch & Dennis J. Devine, *Effects of Reasonable Doubt Definition & Inclusion of a Lesser Charge on Jury Verdicts*, 23 LAW & HUM. BEHAV. 653, 667 (1999). Koch and Devine's data is slightly more complicated than it is described in the text, for they were, at the same time, testing the effect of the inclusion of a lesser offense charge. They found that the Federal Judicial Center instruction led to fewer convictions when there was a lesser included offense charge, but higher when there was not such a charge. *Id.*

[345] *Id.* at 659.

[346] 685 A.2d 1242, 1251-52 (N.J. 1996). The new instruction reads:

> In criminal cases, the government's proof must be . . . beyond a reasonable doubt. A reasonable doubt is an honest and reasonable uncertainty in your minds about the guilt of the defendant after you have given full and impartial consideration to all of the evidence. A reasonable doubt may arise from the evidence itself or from a lack of evidence. It is a doubt that a reasonable person hearing the same evidence would have.
>
> Proof beyond a reasonable doubt is proof, for example, that leaves you firmly

the conviction rate in New Jersey after *Medina* did not show any statistically significant change.[347] He notes, though, that the total number of cases increased, while the percentage of cases tried declined, a dichotomy that he believes does not provide "disconfirming evidence" that the revised instruction is a greater constraint on juror behavior.[348] Solan correctly notes that there are far too many variables at play to reach strong conclusions based on this data.[349] The absence of strong evidence suggests to me, however, the weak effect that the change made by the New Jersey Supreme Court has had. The point is not that the change did not have some effect; I imagine that in a couple of cases, the change might have made a difference. But in the vast majority of criminal cases in New Jersey, the change in the reasonable doubt instruction had *no* effect on how jurors behaved.

The existing evidence, therefore, suggests that alternatives to the existing reasonable doubt instruction will have some ability to push the standard of proof higher, but not much. There is all the more reason to be skeptical about the effect of changes to the instruction when we consider the evidence that jurors' evaluation of the evidence may also be affected by alterations in the jury instructions. Nagel and Neef note that where subjects were forced, through judicial instructions, to raise *SP* from .60 to .90, the subjects also raised their subjective evaluation of the probability that the defendant was guilty; similarly, subjects lowered their evaluations of the probability of the defendant's guilt in response to instructions to lower *SP*.[350] If this is true, then there is good reason to suspect that even where jury instructions are effective at raising the value that jurors will give to *SP*, the effect will be diminished or even be eliminated by the jurors' reweighing of the evidence against the defendant. The result is that jury instructions may have little effect on jurors' decisions on whether to convict or acquit.

> convinced of the defendant's guilt. In this world, we know very few things with absolute certainty. In criminal cases the law does not require proof that overcomes every possible doubt. If, based on your consideration of the evidence, you are firmly convinced that the defendant is guilty of the crime charged, you must find him guilty. If, on the other hand, you are not firmly convinced of defendant's guilt, you must give defendant the benefit of the doubt and find him not guilty.

[347] *See* Solan, *supra* note 9, at 129.

[348] *Id.* at 130.

[349] *Id.* Indeed, there is another way to explain the data. It is possible that as the number of cases rose, the total amount of judicial and prosecutorial resources stayed constant, which then necessitated a lower number of trials.

[350] NAGEL & NEEF, *supra* note 59, at 198; Nagel et al., *supra* note 56, at 371.

Finally, even if changes in the instruction have some effect, the resulting upward effect may be offset by the costs that result from disruption to the existing system. When change in the reasonable doubt instruction result from appellate cases, such as *Medina* and *Cage v. Louisiana*,[351] the short-term effect can be numerous direct and collateral challenges to prior convictions. Even if the appropriate court concludes that the rule should not be applied retroactively, the resulting litigation may be particularly costly. If there is little benefit to accrue from a change in the instruction, the offsetting subsequent litigation costs may be more than enough to counsel against the change.

CONCLUSION

At the end of every American criminal jury trial, the judge tells the jurors that they may convict only if they are convinced of the defendant's guilt beyond a reasonable doubt. The meaning of these words has eluded judges and juries for nearly 200 years. The result is that jurors are left to decide for themselves how much certainty to require in any given case. Because jurors are at least somewhat rational, decisions about where to set the standard of proof are likely to vary from case to case, depending upon the costs and benefits of accurate and inaccurate verdicts in a particular case.

This Article has suggested that the variation in the standard of proof that results from this state of affairs is generally preferable, because it assures that jury decisions will generally mimic the standard of proof that society would apply in a particular case. This presumes, of course, that society, or its democratic representative (the jury), is the proper body to be making such decisions. While that point is certainly debatable, in the absence of any other clear way to reach normative decisions about the proper standard of proof, the best we can hope for is to replicate society's decisions.

None of this means that the standard is, or should be, low, at least in any serious criminal case. To the contrary, in many cases the standard of proof *should* be high because of the high costs of erroneous convictions in relation to the other benefits. Of course, there are cases where it should be higher than others. For instance, in death penalty cases, we would presumably want the standard of proof to be even higher than in other cases because of the even larger costs that flow from erroneous convictions in such cases. Similarly, we can imagine that there are cases

---

[351] 498 U.S. 39 (1990) (per curiam).

where the standard should be lower. In cases where we are dealing with low-level offenses, particularly ones that do not involve jail time, the standard perhaps should be lower because the costs associated with erroneous convictions are lower. Similarly, cases involving repeat offenders perhaps should also have a lower standard of proof because the relative benefits of convictions, and the relative costs of acquittals, are higher than in other cases.

Many people will disagree with some or all of the conclusions in the preceding paragraph about how to vary the standard of proof. But that in a way is the point: it is quite likely that many people in our society would reach very different conclusions about how to adjust the reasonable doubt standard. The key to the reasonable doubt instruction is that it permits individual juries to wrestle with this decision and does not tell them how to resolve it. Although the instruction encourages jurors to apply a high standard of proof, it permits them (through omission) to apply a somewhat lower one, at least in a case where they agree such a standard is appropriate.

This flexibility also goes a long way to explaining why the reasonable doubt standard has been so stable over the last 100 years. The standard, as applied, can be either higher or lower in response to circumstances. When more convictions are needed, the standard will drop; when less are needed, the standard will rise. The consequence is that the resulting applied standard is rarely at odds with society's criminal justice needs, and there is little desire to change it.

As a result, reasonable doubt is likely to remain a cornerstone of American criminal justice, at least so long as juries are at the center of its trials. But this does not mean that we should be complacent about how the reasonable doubt standard functions. There may be many ways in which one side (or both) can manipulate the trial to alter the way in which the standard of proof operates. I have pointed out that present death penalty jury selection rules may have a negative effect on the applied standard of proof. The result is likely too many erroneous convictions in death penalty cases. Other rules, particularly those governing the admissibility of evidence, may manipulate the standard of proof by skewing the information the jury receives. Indeed, this Article implicitly suggests that giving jurors *more* information about crimes, defendants, and punishments is good, a conclusion that is contrary to the general trend in academic thinking about such matters. Recognizing that the reasonable doubt standard invites juries to consider matters that are technically "irrelevant" may help explain and justify rules that allow the admission of such evidence. Enriching our understanding of reasonable

doubt can therefore expand how we approach related issues in criminal law and procedure.

APPENDIX

For *SP*(rape) = *SP*(theft), you must have

$(U_{cg}(rape) - U_{ag}(rape))/(U_{ai}(rape) - U_{ci}(rape)) =$
$(U_{cg}(theft) - U_{ag}(theft))/(U_{ai}(theft) - U_{ci}(theft))$.

Because $U_{cg}(rape) > U_{cg}(theft)$ and $U_{ag}(theft) > U_{ag}(rape)$, it follows that $U_{cg}(rape) - U_{ag}(rape) > U_{cg}(theft) - U_{ag}(theft)$.

Then we know that $U_{cg}(rape) - U_{cg}(theft) > 0$ and $U_{ag}(theft) - U_{ag}(rape) > 0$.

This proposition, however, is almost certainly false.

If $U_{cg}(rape) - U_{ag}(rape) = U_{cg}(theft) - U_{ag}(theft)$, then $U_{cg}(rape) - U_{cg}(theft) = U_{ag}(rape) - U_{ag}(theft)$, which implies that $U_{ag}(rape) - U_{ag}(theft) > 0$, which would contradict $U_{ag}(theft) - U_{ag}(rape) > 0$.

Similarly, if $U_{cg}(rape) - U_{ag}(rape) < U_{cg}(theft) - U_{ag}(theft)$, that would imply that then $U_{cg}(rape) - U_{cg}(theft) < U_{ag}(rape) - U_{ag}(theft)$, which would imply that $U_{ag}(rape) - U_{ag}(theft) > 0$, because $U_{cg}(rape) - U_{cg}(theft) > 0$. But given $U_{ag}(theft) - U_{ag}(rape) > 0$, this is not possible.

Similarly, if we are assuming that $U_{ai}(rape) = U_{ai}(theft)$, then

$U_{ai}(rape) - U_{ci}(rape) < U_{ai}(theft) - U_{ci}(theft)$.

This is because if we assume that $U_{ai}(rape) = U_{ai}(theft)$, then the equation $U_{ai}(rape) - U_{ci}(rape) > U_{ai}(theft) - U_{ci}(theft)$ would imply that

$U_{ai}(rape) - U_{ai}(theft) > U_{ci}(rape) - U_{ci}(theft)$,

This equation can be reduced to $0 > U_{ci}(rape) - U_{ci}(theft)$, which is impossible given our assumption that $U_{ci}(rape) > U_{ci}(theft)$.

***