

The Case for Varying Standards of Proof

GUSTAVO RIBEIRO*

TABLE OF CONTENTS

I.	INTRODUCTION	162
II.	BACKGROUND: THE ERROR-DISTRIBUTION FUNCTION OF STANDARDS OF PROOF	165
III.	A PROBLEM AND A PROPOSAL	174
IV.	NORMATIVE ARGUMENTS IN FAVOR.....	183
	A. <i>Welfare Considerations</i>	183
	B. <i>Fairness Considerations</i>	189
	C. <i>Distributive Considerations</i>	194
V.	A POSITIVE ARGUMENT IN FAVOR.....	198
VI.	OBJECTIONS AGAINST AND REPLIES	207
	A. <i>High Administrative Costs and Unpredictable Effects</i>	207
	B. <i>Cognitive Limitations</i>	212
	C. <i>Already Varying De Facto Standards of Proof</i>	216
VII.	CONCLUSION	219

* © 2019 Gustavo Ribeiro. Boston University School of Law. Thank you to Ronald J. Allen, Scott Brewer, Ali Butler, Hasan Dindjer, Rachel Herdy, Louis Kaplow, Gary Lawson, Nadav Orian Peer, Andrea J. Schweitzer, Alex Stein, Alex Whiting, and William Twining for their helpful comments and suggestions on previous drafts and to the *San Diego Law Review* team for the terrific editorial work. For useful discussions, I am grateful to participants at the Byse Workshop, the S.J.D. Colloquium, and the Law and Philosophy Society at Harvard Law School, the Jurisprudence Discussion Group at the University of Oxford, and the Legal Theory Seminar Series at the University of Edinburgh.

I. INTRODUCTION

The current evidence rules group cases that ought to remain separate for a myriad of reasons. Most dramatically, we assign the same standard of proof to drastically different cases under the justification that we accept, or should accept, the same error-distribution for those cases.¹ To justify the use of the “preponderance of the evidence” standard in most civil litigation cases, courts and scholars frequently repeat that—in the majority of civil cases—society is, or should be, indifferent to errors in favor of plaintiffs or defendants.² The inescapability of the beyond “reasonable-doubt standard” in criminal law is defended as embodying a constant defendant-friendly error-distribution across all criminal adjudication.³ However ubiquitous, these justifications are twice mistaken.

First, they are normatively mistaken. There are important welfare, fairness, and distributive considerations in favor of varying standards of proof across different types of cases. Analyses based on welfare considerations tell us that if the benefit from deterring harmful acts is greater than the loss from chilling benign acts, then we have a strong argument that reducing the standard of proof might be socially desirable; whereas if the reverse is true, an increase might be preferable.⁴ Under this view, the socially optimal

1. See, e.g., *Santosky v. Kramer*, 455 U.S. 745, 745–46 (1982) (“In any given proceeding, the minimum standard of proof tolerated by the due process requirement reflects not only the weight of the public and private interests affected, but also a societal judgment about how the risk of error should be distributed between the litigants.”); *Addington v. Texas*, 441 U.S. 418, 423 (1979) (“The standard serves to allocate the risk of error between the litigants and to indicate the relative importance attached to the ultimate decision.”); *In re Winship*, 397 U.S. 358, 369–70 (1970) (Harlan, J., concurring) (“[E]ven though the labels used for alternative standards of proof are vague and not a very sure guide to decisionmaking, the choice of the standard for a particular variety of adjudication [reflects] a very fundamental assessment of the comparative social costs of erroneous factual determinations.”).

2. *Addington v. Texas*, 441 U.S. 418, 423 (1979) (“Since society has a minimal concern with the outcome of [monetary disputes between private parties,] plaintiff’s burden of proof is a mere preponderance of the evidence. The litigants thus share the risk of error in roughly equal fashion.”).

3. *Winship*, 397 U.S. at 363 (“The reasonable-doubt standard plays a vital role in the American scheme of criminal procedure. It is a prime instrument for reducing the risk of convictions resting on factual error. The standard provides concrete substance for the presumption of innocence—that bedrock ‘axiomatic and elementary’ principle whose ‘enforcement lies at the foundation of the administration of our criminal law.’” (quoting *Coffin v. United States*, 165 U.S. 432, 453 (1895))).

4. See, e.g., Louis Kaplow, *Burden of Proof*, 121 *YALE L.J.* 738, 855–56 (2012) [hereinafter Kaplow, *Burden of Proof*]; sources cited *infra* note 81. Alternatively, for a discussion against the idea of a monolithically *social* and an argument that there are only competing interests, values, and preferences, in constant combat, differentially aided by law and legal processes, see, for example, Oliver Wendell Holmes, Jr., *Privilege, Malice*,

standard occurs where these two opposing tendencies—deterrence of harmful acts, on the one hand, and chilling of benign acts, on the other—offset each other. Most importantly, we have good reason to believe these two forces vary drastically in different cases. This suggests the optimal standard also varies with different types of cases.

Taking fairness as the guiding consideration, there is a strong argument that people have a right to procedures that assign the appropriate importance to the risk of being unjustly found liable or guilty.⁵ If this risk varies in different cases because, for instance, we value the injustices that follow from wrongful convictions or acquittals differently—such as capital punishment versus minor misdemeanors—it should follow that people have a right in different cases to different legal procedures that weigh the risk of an unjust verdict. Standards of proof are perhaps the most salient example of legal procedures with that effect.⁶

Distributive considerations push us to understand the levels of the standards of proof as part of people's background legal entitlements that influence the societal distribution of wealth and income.⁷ This suggests we should not ignore the disparate effects on the distribution of wealth and income that a system that assigns the same standard of proof across the board has on differently situated plaintiffs and defendants. We also gain an important insight for programmatic projects. Legal reforms intended to alter the distribution of wealth or income among different social groups can utilize standards of proof to reach their objectives.

Second, the justifications for grouping different cases under the same standard of proof is positively mistaken. Research on jury decision-making suggests that, under certain circumstances, jurors tend to reduce the de facto standard of proof necessary for conviction.⁸ We should adjust our

and Intent, 8 HARV. L. REV. 1, 3 (1894); Karl N. Llewellyn, *A Realistic Jurisprudence—The Next Step*, 30 COLUM. L. REV. 431, 461–62 (1930).

5. See RONALD DWORKIN, *A MATTER OF PRINCIPLE* 3, 119–21 (1985); HO HOCK LAI, *A PHILOSOPHY OF EVIDENCE LAW: JUSTICE IN THE SEARCH FOR TRUTH* 213, 223–24 (2008).

6. See DWORKIN, *supra* note 5, at 84–88, 92–93.

7. See, e.g., Duncan Kennedy, *Distributive and Paternalist Motives in Contract and Tort Law, with Special References to Compulsory Terms and Unequal Bargaining Power*, 41 MD. L. REV. 563, 563, 565 (1982) [hereinafter Kennedy, *Distributive and Paternalist*]; Duncan Kennedy, *Cost-Benefit Analysis of Entitlement Problems: A Critique*, 33 STAN. L. REV. 387, 422–23 (1981) [hereinafter Kennedy, *Cost-Benefit Analysis*].

8. See, e.g., DENNIS J. DEVINE, *JURY DECISION MAKING: THE STATE OF THE SCIENCE* 39–40 (2012); REID HASTIE ET AL., *INSIDE THE JURY* 231 (1983).

discourse to match reality. Not only is a system with greater variation in the number of standards of proof normatively desirable but it is also consistent with how jurors decide cases. Just like we weigh different circumstances and cases through sanctions, liability, and procedural and evidentiary rules, we can, and should, also express these differences by using different standards of proof in different types of cases.

This proposal will undoubtedly face objections. One might be concerned about the potentially high administrative costs associated with a system of varying standards of proof.⁹ Although an important concern for any policy proposal, we should be careful not to overestimate this point. A related objection pertains to the difficulty of predicting the dynamic effects of a system with varying standards of proof.¹⁰ Once a new standard is in place, people will modify their behavior to conform to the new standard. As such, we would need to predict avoidance and exploitation efforts and continuously adjust our standards. But these continuous adjustments would require an immense amount of information. For objectors, the twin problems of high costs and unpredictable dynamic effects will most likely offset any potential gain resulting from a system with varying standards of proof. These two objections touch on a more general point about the function of standards serving as mechanisms for distributing errors.

To make sure a specific distribution of errors actually occurs in our society, we need information concerning, among other things, “the distribution of truly guilty and truly innocent [people] who go to trial.”¹¹ This information, however, is prohibitively costly, or impossible, to obtain. One way to change our understanding of how standards can satisfactorily distribute errors is by lowering our expectations about our capacity to verify whether we have reached an appropriate error-distribution. We can also see standards of proof as prescriptive generalizations with a relevant underlying justification—namely to achieve a given error distribution. Similar to legal rules, standards of proof are suboptimal generalizations. But unlike legal rules, this suboptimality does not come from over- and

9. See KEVIN M. CLERMONT, STANDARDS OF DECISION IN LAW: PSYCHOLOGICAL AND LOGICAL BASES FOR THE STANDARD OF PROOF, HERE AND ABROAD 264–68 (2013); Ronald J. Allen & Alex Stein, *Evidence, Probability, and the Burden of Proof*, 55 ARIZ. L. REV. 557, 580–81 (2013).

10. See Kaplow, *Burden of Proof*, *supra* note 4, at 809, 855.

11. LARRY LAUDAN, TRUTH, ERROR, AND CRIMINAL LAW: AN ESSAY IN LEGAL EPISTEMOLOGY 73 (2006) (citing Ronald 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 n.65 (1997) [hereinafter Allen, *Winship*]); Ronald J. Allen, *Rationality, Algorithms and Juridical Proof: A Preliminary Inquiry*, 1 INT’L J. EVIDENCE & PROOF 254, 260 (1997) [hereinafter Allen, *Juridical Proof*]; Michael S. Pardo, *Second-Order Proof Rules*, 61 FLA. L. REV. 1083, 1088 (2009).

under-inclusiveness. Rather, it comes from constraints on our ability to verify whether the targeted error distribution has been achieved.

Another objection I address is the claim that the legal system already adjusts the distribution of error across different types of cases through other legal mechanisms, such as adding or removing causes of action for certain subsets of cases. Insofar as these other instruments impact how easy it is to impose liability on potential defendants, they can cause the de facto standard of proof to vary. After considering the effects of altering the de facto standard using a procedural or substantive strategy, I conclude that the procedural strategy is either superior or equally preferable to the substantive strategy, which gives us a good reason to prefer the former. This suggests it will often be best to alter the error-distribution to a socially desirable ratio through procedural instruments.

This Article is divided into five additional parts. Part II discusses the two functions usually attributed to standards of proof. Special focus is placed on the standards serving as a mechanism to distribute the risk of factual error between plaintiffs and defendants. Part III suggests that once we seriously consider this function of standards, we can conclude that we group cases under a restricted number of standards. This Article argues for a system with a greater number of standards of proof that inform legal decision-making. Part IV considers welfare, fairness, and distributive arguments in favor of varying standards of proof. Part V presents a positive argument in favor of the proposal. Lastly, Part VI responds to potential objections.

II. BACKGROUND: THE ERROR-DISTRIBUTION FUNCTION OF STANDARDS OF PROOF

Standards of proof are a tricky evidentiary mechanism. Although they are ubiquitous in legal decision-making, there is widespread disagreement about how exactly to understand them.¹² This lack of understanding greatly

12. Empirical research suggests considerable confusion regarding standards exists in practice, both with respect to articulation and communication of the standards of proof by trial judges and to comprehension and application by the juries. *See, e.g.*, Joel D. Lieberman, *The Psychology of the Jury Instruction Process*, in 1 *PSYCHOLOGY IN THE COURTROOM: JURY PSYCHOLOGY* 129, 129, 132, 139–40 (Daniel A. Krauss & Joel D. Lieberman eds., 2009) (reviewing the existing literature); *see also* CLERMONT, *supra* note 9, at 113; Dorothy K. Kagehiro, *Defining the Standard of Proof in Jury Instructions*, 1 *PSYCHOL. SCI.* 194, 196, 198 (1990); Federico Picinali, *Is “Proof Beyond a Reasonable Doubt” a Self-Evident Concept? Considering the U.S. and the Italian Legal Cultures*

affects our capacity to articulate and communicate standards in a precise and defensible manner. Existing formulations are often described as obscure, incoherent, or worse.¹³ This disagreement, in turn, reveals a lack of familiarity with the functions standards are meant to serve.

Standards of proof are thought to serve essentially two related functions.¹⁴ First, they are decision-making thresholds—standards tell us whether the available evidence in favor or against a particular evidentiary hypothesis offers enough evidential support for us to take a given proposition as proven for some particular purpose.¹⁵ Consider medical drug testing as an example.¹⁶ Scientists and public health policymakers face a difficult tradeoff when deciding whether to release drugs to the public. On the one hand, they want to make drugs publicly available as quickly as possible to prevent the spread of diseases. On the other hand, they want to be confident that the drug is effective and does not cause severe side effects. The problem is that, to be sure of a drug's effectiveness and safety, scientists and policymakers must expend time and resources that might not be available.

To resolve this tension between speed and safety, scientists and policymakers usually establish a decision-making threshold.¹⁷ If the available data offers some level of support for the proposition that the drug is effective and safe, then they are willing to accept the proposition that

Towards the Understanding of the Standard of Persuasion in Criminal Cases, GLOBAL JURIST, 2009, at 1, 11, 26 (finding empirical studies that show widespread misunderstanding of the importance of *beyond a reasonable doubt*); Frederick Schauer, *Slippery Slopes*, 99 HARV. L. REV. 361, 370–73 (1985) (discussing the challenges involved in formulating and communicating precise linguistic formulations of legal principles); David U. Strawn & Raymond W. Buchanan, *Jury Confusion: A Threat to Justice*, 59 JUDICATURE 478, 478–80 (1976) (suggesting that after instructions only 50% of jurors understand that the defendant did not have to present any evidence of his innocence and that the state had to establish the defendant's guilt beyond reasonable doubt).

13. See LAUDAN, *supra* note 11, at 61, 64.

14. The discussions that follow are premised on the idea that the applicable standard of proof has significant implications for how the guilty and the innocent fare in the judicial system. Given that an overwhelming majority of cases settle before trial, see, for example, Daniel Epps, *The Consequences of Error in Criminal Justice*, 128 HARV. L. REV. 1065, 1114 (2015), one might wonder whether standards are merely a sideshow. One well-known answer is to contend that standards are still important because settlements and pleas occur in the shadow of trial outcomes. See Stephanos Bibas, *Plea Bargaining Outside the Shadow of Trial*, 117 HARV. L. REV. 2463, 2464–66 (2004); Oren Gazal-Ayal & Avishalom Tor, *The Innocence Effect*, 62 DUKE L.J. 339, 383–86, 388–96 (2012); Alec Walen, *Proof Beyond a Reasonable Doubt: A Balanced Retributive Account*, 76 LA. L. REV. 355, 401–02 (2015). This article proceeds on this assumption.

15. See LAUDAN, *supra* note 11, at 64.

16. See, e.g., MELISSA LEACH & JAMES FAIRHEAD, *VACCINE ANXIETIES: GLOBAL SCIENCE, CHILD HEALTH AND SOCIETY* 17 (2007).

17. See *id.*

the drug is effective and safe.¹⁸ In other words, scientists and policymakers establish a minimal threshold of evidential support that the evidence must reach to consider a given proposition as proven. Essentially, they establish a standard of proof.

A second, related, function of standards of proof courts and scholars often highlight—and the one I focus on—is that of distributing errors.¹⁹ To understand this different, yet connected, function let us start with a question: Why would anyone want to establish a standard of proof higher than *more probable than not*? After all, if a proposition is probably more true than not, then, under epistemic justification, we are at least *prima facie* rationally justified in believing that proposition.²⁰ Why require something more demanding than what rationality allows?

To answer this question, let us return to the drug testing example. Suppose you oversee a massive governmental drug-development program. You collect data from the most prestigious laboratories, and the results clearly seem to show that drug X is only *probably* more effective and safer than not. Under normal conditions—other than a catastrophic epidemic—would you authorize the drug program based on this evidence alone? I believe most people would not. Most people weigh the cost of the error of releasing a drug that is ineffective or unsafe more heavily than they do the cost of the error of delaying the release of a safe and effective drug.

Weighing different errors differently is not a phenomenon specific to public health situations. Something similar takes place in the legal system. Because we cannot completely eliminate errors from an adjudicative system with limited resources, society must decide which errors are more serious and worth spending more resources to avoid. Standards of proof are the best example of legal mechanisms primarily concerned with this

18. *See id.*

19. *See supra* note 1; *see also* MICHAEL O. FINKELSTEIN, QUANTITATIVE METHODS IN LAW: STUDIES IN THE APPLICATION OF MATHEMATICAL PROBABILITY AND STATISTICS TO LEGAL PROBLEMS 67 (1978); LAUDAN, *supra* note 11, chs. 2–3; ALEX STEIN, FOUNDATIONS OF EVIDENCE LAW 138 (2005) (“[E]vidential rules and principles affiliating to [the Anglo-American systems of evidence] have a single all-important function: allocation of the risk of error.”); Richard S. Bell, *Decision Theory and Due Process: A Critique of the Supreme Court’s Lawmaking for Burdens of Proof*, 78 J. CRIM. L. & CRIMINOLOGY 557, 563, 579–81 (1987); Michael Tigar, *The Supreme Court, 1969 Term*, 84 HARV. L. REV. 1, 158 n.13 (1970) (“[T]he reasonable doubt standard seeks to assure that erroneous acquittals of the guilty are far more common than the erroneous convictions of the innocent.”).

20. *See generally, e.g.*, EARL CONEE & RICHARD FELDMAN, EVIDENTIALISM: ESSAYS IN EPISTEMOLOGY (2004).

task. If we could compute the relative costs and benefits of all accurate and erroneous convictions and acquittals in every decision, we might be able to use this cost ratio to determine the height of the standard.²¹

One way to visualize this function of standards is through graphs like the one depicted in Figure 1 below.²² Such illustrations show the relations between the standard of proof and the distribution of the types of errors we expect to occur. The horizontal axis represents a spectrum of probabilities that fact-finders, either judges or jurors, assign to defendants' "apparent guilt." At one end are cases that tend to go to trial with very weak evidence—for instance, cases involving private, solo conduct with low externalities and low probability of detection. At the other end are cases that tend to go to trial with very strong evidence against defendants, such as public daylight conduct with many witnesses, externalities, and high detection rates. The vertical axis is the relative frequency or proportion of cases that exhibit apparent guilt.²³

We are interested here in two distributions: one representing truly innocent defendants and another representing truly guilty defendants.²⁴ Each distribution represents the likely values of apparent guilt for the two types of defendants.²⁵ It is important to clarify at the outset a few important assumptions. Absent a highly dysfunctional legal system, we should expect that truly guilty defendants will, on average, have higher apparent guilt than truly innocent defendants. As a consequence, the distribution of truly guilty defendants is represented to the right of the truly innocent distribution on the horizontal spectrum. However, the exact location of each distribution

21. LAUDAN, *supra* note 11, at 68.

22. See *infra* p. 169. For examples of other works using similar diagrams to represent the relation between standards of proof and the distribution of the risk of error between parties, see LAUDAN, *supra* note 11, 66–70; Ronald J. Allen, *Burdens of Proof*, 13 LAW, PROBABILITY & RISK 195, 201, 204–06, 211–12 (2014); Bell, *supra* note 19, at 570–82; Michael L. DeKay, *The Difference Between Blackstone-Like Error Ratios and Probabilistic Standards of Proof*, 21 LAW & SOC. INQUIRY 95, 101, 105, 113, 120, 123–25 (1996).

23. This makes the area under each curve equal to 1—representing 100% of the data—and explains why both distributions are shown with similar sizes.

24. Some might object to the use of the terms *guilty* and *innocent*. Under this view, guilty and innocent represent ascriptions that only come at the end of a specifically legal procedure. See, e.g., Zenon Bankowski, *The Value of Truth: Fact Scepticism Revisited*, 1 LEGAL STUD. 257, 265 (1981). To not confuse readers familiar with the literature, I use these terms simply to adopt the same terminology used by other works on this topic. One could use easily alter the names of both curves without harm to the lessons that follow.

25. Another way to understand these distributions is as conditional distributions reflecting the likely values of apparent guilt—given that the defendant is truly innocent or truly guilty. DeKay, *supra* note 22, at 101–02. We can then derive the unconditional probabilities—for instance, that a randomly selected defendant is acquitted and truly innocent—from the conditional probabilities. See *id.* All one must do is to adjust the conditional probabilities to reflect the mix of truly guilty and innocent defendants appearing before the court. See *id.*

along the x-axis is not crucial for our purposes. Also, although the graph depicts both distributions using normal curves, normality is not an essential feature of this analysis. “Other unimodal distributions [can] yield similar results”²⁶ Another important assumption is that the distributions overlap. This is because of significant variability within each group. Some truly innocent defendants have a higher apparent guilt than some truly guilty defendants.

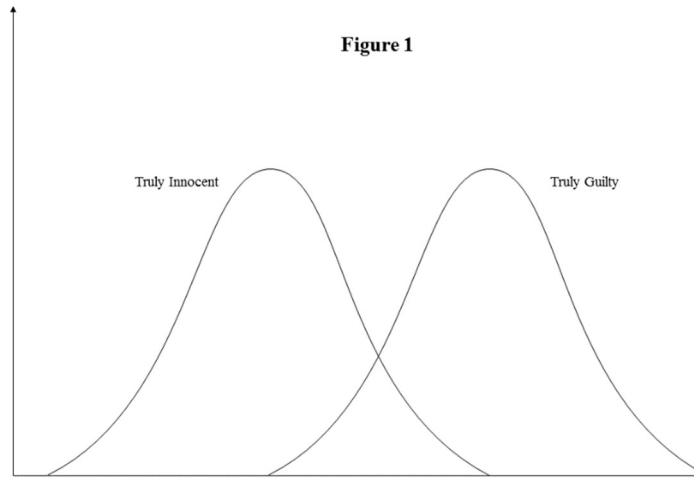
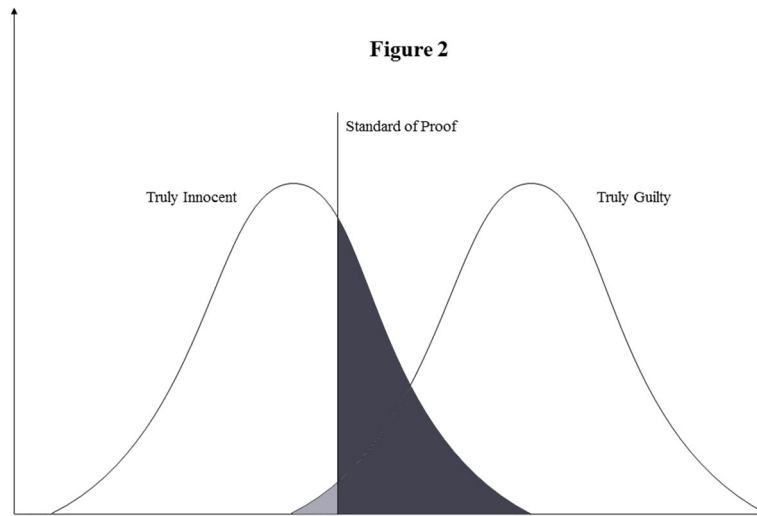


Figure 1 does not show how many findings will be correct or mistaken, nor does it show how the errors will be distributed. We can illustrate setting up a standard of proof by drawing a line perpendicular to the horizontal axis. That line marks how strong the admitted evidence supporting a conclusion about liability must be to legally justify a conviction or finding of liability. More specifically, it shows how strong the admitted evidence supporting the case of the party carrying the burden of persuasion must be. Figure 2 represents a possibility.

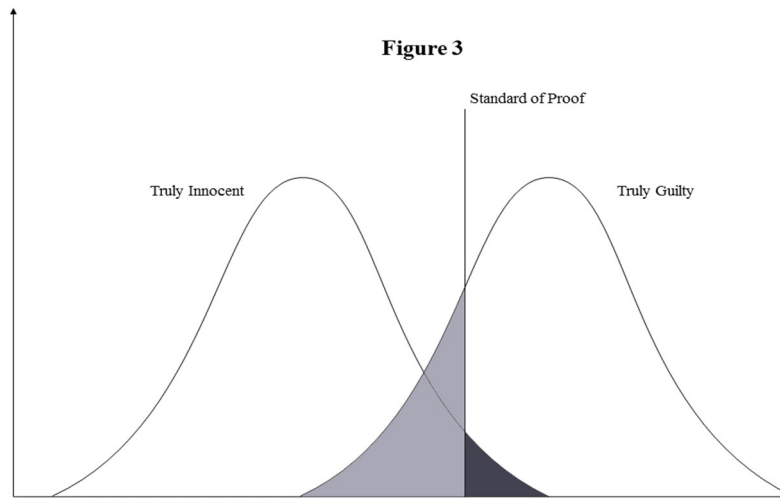
26. *Id.* at 102.



This mode of representation is visually appealing because the areas of these regions are closely related to the probabilities and frequencies of the four possible outcomes of a verdict. The area under the “Truly Guilty” curve and to the right of the standard of proof line represents the probability that a truly guilty individual will be convicted at trial. The area under the “Truly Guilty” curve to the left of the standard—the small light gray area—indicates the probability of a truly guilty individual being mistakenly acquitted at trial. Conversely, the probability that a truly innocent individual goes to trial and is mistakenly convicted is shown by the larger dark gray area under the “Truly Innocent” curve to the right of the standard line. For all other values to the left of that point, the individual is correctly found not liable.

Presumably, we want to minimize the likelihood that a truly innocent individual is mistakenly found liable. Holding other features of the legal system constant, such as level of enforcement, sanctions, *et cetera*, one way we can attempt to reduce mistaken liability verdicts is by increasing the minimal level of evidential support required for a liability conclusion. Graphically, this means moving Figure 2’s standard of proof line further to the right. This would reduce the dark gray area under the “Truly Innocent” curve, translating to a reduction in the probability of truly innocent defendants mistakenly found liable.²⁷ Figure 3 illustrates this.

27. This point seems intuitively correct. Raising the standard of proof will likely increase the percentage of truly guilty individuals that are convicted out of the entire set



Note, however, that with the same stroke we would also reduce the likelihood of correctly assigning liability to a truly guilty individual. In other words, when determining the optimal standard, the errors of false positives versus the errors of false negatives are traded off. The greater the likelihood that we will correctly assign liability to truly guilty individuals,

of convicted individuals. This is particularly true if we hold constant the frequency with which cases come before the court—keeping the “Truly Innocent” and “Truly Guilty” curves intact as Figures 2 and 3 illustrate. However, depending on the empirical assumptions we start with, we might also expect the exact opposite effect. The problem is that there is a good dose of endogeneity in this kind of analysis—the number of innocent and guilty individuals that go to trial seems to be a function of, among other things, the level of the standard of proof. With a higher standard of proof, a conviction is more difficult, which lowers the expected costs of sanctions for individuals—some of whom might become more likely to commit the act that brings them to court, mistakenly or not. The question then becomes whether we have reason to believe we will see a corresponding increase in the frequency of cases that come before the court. If more cases happen to go to trial as a result of an increase in the standard of proof, this might *reduce* the chances that a higher percentage of individuals convicted are truly guilty, the *opposite* of what seems intuitively correct. The situation, then, seems to be as follows: On the one hand, we have the effect that an increase in the standard of proof increases the chances that a higher percentage of individuals convicted are truly guilty. On the other hand, the fact that we should expect the frequency for both cases to increase might reduce this likelihood. The problem is that these two effects might run in opposite directions and if so, we would have a hard time knowing which effect is stronger and thus more likely to prevail. See Kaplow, *Burden of Proof*, *supra* note 4, at 789–805.

the greater the likelihood that we will mistakenly assign liability to truly innocent individuals. On the flipside, lowering the standard of proof—moving the line in Figures 2 or 3 to the left—increases the likelihood that fact-finders will correctly assign liability to truly guilty individuals while concurrently increasing the likelihood that truly innocent individuals will mistakenly be found liable. In either direction, there is a trade-off.²⁸

If we value one type of error much more than the other, why not simply set the standard at the highest level possible to eliminate virtually all chances of convicting an innocent? The main problem with this proposal is that it ignores the costs associated with both types of errors. If the standard is set too high, we risk making it nearly impossible to convict not only the truly innocent, but also the truly guilty. Setting too high a standard has nontrivial consequences, including: (1) losses in deterrence, potentially causing an increase in crime rates; (2) losses in feelings of justice, because a wrong that was committed is not remedied; and (3) losses in trust in the legal system, because victims are denied retribution.²⁹ This important point is often overlooked. Legal scholarship and practice tend to concentrate solely on the costs following wrongful convictions. Consequently, it has ignored—or grossly underestimated—the costs following wrongful acquittals. These costs are real, and there is certainly a limit to how much we will accept to prevent the flipside of convicting an innocent. Once society concludes these costs are too high, it can adjust the applicable standard of proof to bring about a more desirable tradeoff. This is ultimately a matter of public policy and choices must inevitably be made. There are no pre-determined socially desirable ratios.³⁰

28. This trade-off is partially the result of a significant variability within each group. Some truly innocent defendants appear guiltier than some truly guilty defendants. DeKay, *supra* note 22, at 101. This is visually represented by the overlap of the two curves. Because appearance of guilt is imperfectly related to true guilt, fallible decision-makers will invariably fail to correctly identify all innocent or guilty defendants. *See id.* The level of variability between both groups—the overlap size—has great importance. Anything that makes truly guilty defendants appear guiltier or makes innocent defendants appear less guilty moves the curves apart and increases accuracy. *See, e.g.,* DeKay, *supra* note 22, at 101–02.

29. *See* Ronald J. Allen & Larry Laudan, *Deadly Dilemmas*, 41 TEX. TECH L. REV. 65, 68, 73 (2008); *see also* Ronald J. Allen & Larry Laudan, *Deadly Dilemmas II: Bail and Crime*, 85 CHI.-KENT L. REV. 23, 25–27 (2010); Ronald J. Allen & Larry Laudan, *Deadly Dilemmas III: Some Kind Words for Preventive Detention*, 101 J. CRIM. L. & CRIMINOLOGY 781, 801–02 (2011).

30. Even after including the costs of wrongful acquittal, the picture is still incomplete. It is not enough to weigh only the costs of erroneous decisions, whether convictions or acquittals. *Accurate* convictions and acquittals must also be weighed in reaching any decision about the proper standard of proof. *See, e.g.,* Ronald J. Allen, *How to Think About Errors, Costs, and Their Allocation*, 64 FLA. L. REV. 885, 889 (2012); DeKay, *supra* note 22, at 95; Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal*

The error-distribution function of standards should not be overstated, however. That standards of proof serve as mechanisms for distributing errors is not to say standards are the only force at play in determining the error-distribution. The relationship between standards and the distribution of different types of errors is actually more complicated than it first seems. The precise allocation of errors also depends on factors other than the standard, including the frequency of truly guilty and truly innocent that go to trial and the probability distribution of the cases that get to trial where the admitted evidence is stronger than the applicable standard.³¹ In other words, the error-distribution also depends on how the “Truly Innocent” and “Truly Guilty” curves look.³²

One last point before getting into the arguments for varying standards of proof. It is commonly thought that the function of legal fact-finding is primarily an epistemic one of *getting the facts right*.³³ But after studying evidence, one quickly notes that the epistemic function of fact-finding cannot be the whole story. Still, accuracy remains a prime goal of the evidentiary system. Therefore, a worthy question is: How does the error distribution function of standards connect to the epistemic function of fact-finding?

One hypothesis is that the level of the standard is directly correlated to the level of accuracy in fact-finding.³⁴ So, if we raise the former, we also raise the latter. No matter how intuitive this hypothesis might sound, it is not quite right. The confusion results mostly because certain changes in the legal system can affect *both* accuracy and standards. Consider the effects of providing additional resources for indigent defendants’ counsels, or similarly reducing their caseload. One expected result is improving the overall

Process, 84 HARV. L. REV. 1329, 1381–82 (1971). We should set standards with our eyes on the relative costs and benefits of both accurate and erroneous convictions and acquittals.

31. See Allen, *Winship*, *supra* note 11, at 47 n.65 (“Without 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.”). For concrete examples, see FINKELSTEIN, *supra* note 19, at 59–60. But see Bell, *supra* note 19, at 574–75 (conveying different results concerning preponderance of the evidence, but no clear and convincing evidence and beyond a reasonable doubt).

32. See *supra* Figures 1–3. For a more in-depth examination, see *infra* Section VI.A.

33. See *Tehan v. United States*, 382 U.S. 406, 416 (1966) (“The basic purpose of a trial is the determination of truth . . .”).

34. For discussions about the value of accuracy in fact-finding, see, for example, Kaplow, *Burden of Proof*, *supra* note 4, at 741, 746; Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307, 307–08 (1994) [hereinafter Kaplow, *Accuracy in Adjudication*]; Jack B. Weinstein, *Some Difficulties in Devising Rules for Determining Truth in Judicial Trials*, 66 COLUM. L. REV. 223, 228–29 (1966).

quality of evidence admitted at trial. For example, counsels for indigent defendants will have more time to prepare better defenses and will be able to hire investigators and forensics experts. On the one hand, this might correlate with more accurate fact-finding. Better evidence should provide more support for evidentiary propositions that are more likely to be true. On the other hand, increasing resources for counsels might also cause a de facto increase in the level of the standards of proof.³⁵ Because defendant counsels will be able to prepare better defenses, we should expect that on average it will be harder to convict defendants. This is an example of a change that has the potential to affect both accuracy and the level of standards. We can also envision changes in the legal system that could affect one but not the other. Suppose that in addition to increasing the resources available to indigent defendants' counsels, we simultaneously increase resources available to prosecutors, effectively restoring the prior de facto standard.

These examples indicate the importance of separating accuracy and the level of the standard is not merely conceptual. Although both work together in many instances, we can design changes that would affect one but not the other. This point is important for my purposes. Below, I argue that the legal system should have more flexibility to impose varying standards of proof. This means sometimes increasing, but also sometimes lowering, the applicable standard. Regardless, it should be clear from the outset that lowering the standard does not necessarily mean a decrease in accuracy.

III. A PROBLEM AND A PROPOSAL

This Part argues that there is a problem with the rules currently governing standards of proof. We group cases under a very restricted number of standards. This problem requires a solution. The solution I explore in this Article requires moving in the direction of a system with a greater number of standards of proof to inform legal decision-making.

Civil courts often repeat they do not have a reason to favor error for one party over the other.³⁶ This perceived equality of the relative importance of errors is the driving force behind the preponderance of the evidence standard in almost all civil cases in the United States.³⁷ It is also a common justification for a specific formulation of that standard—namely that the decision-maker should decide if the plaintiff's case is “more likely than

35. Kaplow, *Accuracy in Adjudication*, *supra* note 34, at 356–57.

36. *See, e.g.*, *Addington v. Texas*, 441 U.S. 418, 423 (1979).

37. *See, e.g.*, Ronald J. Allen, *Burdens of Proof, Uncertainty, and Ambiguity in Modern Legal Discourse*, 17 HARV. J.L. & PUB. POL'Y 627, 633–34 (1994); Ronald J. Allen, *The Error of Expected Loss Minimization*, 2 LAW, PROBABILITY & RISK 1, 3 (2003); Pardo, *supra* note 11, at 1088–89; Lawrence B. Solum, *Procedural Justice*, 78 S. CAL. L. REV. 181, 312 (2004).

not,” which is often articulated as having a greater than 50% probability.³⁸ By setting the standard that way, under certain assumptions, the preponderance rule roughly aims to divide the risk of error equally among the parties.³⁹

There are significant exceptions, however. These include certain types of civil lawsuits in which errors favoring plaintiffs are more significant or costly than errors favoring defendants. Consequently, the risk of error is asymmetrically skewed against plaintiffs. In these cases, the “clear and convincing evidence” standard applies to one or more elements of the plaintiff’s claim. Examples include fraud,⁴⁰ civil commitment,⁴¹ deportation,⁴² denaturalization,⁴³ termination of parental rights,⁴⁴ decisions to terminate life,⁴⁵ and proof of malice in defamation cases involving public figures.⁴⁶

However, these exceptions are restricted in number and scope. The general belief is that in civil cases—in which the applicable standard is

38. Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399, 408 (1973). It is important to note that a preponderance of the evidence standard does not mean that only roughly half of the verdicts will be accurate. *Id.* This is because an increase or decrease in either

the plaintiff’s stakes or the effectiveness of his litigation expenditures . . . will induce the plaintiff to spend at a higher rate than the defendant, and vice versa. If the allegations essential to one party’s claim are in fact true, ordinarily it will be easier for him to prove them than for his opponent to disprove them, assuming they spend the same amount of money on the trial.

Id. at 431. “[T]he effectiveness of the expenditures of the party with the meritorious claim will be high relative to the effectiveness of his opponent’s expenditures. This will induce the first party to spend more heavily on litigation than the second.” *Id.* at 431–32. But within the errors, we should expect an equal number of plaintiffs and defendants. *Id.* at 431–35.

39. See, e.g., *Grogan v. Garner*, 498 U.S. 279, 286 (1991) (“[T]he preponderance-of-the-evidence standard results in a roughly equal allocation of the risk of error between litigants . . .”). This is *roughly*—not *exactly*—because the defendant wins if there is a tie. See *id.* One justification for this is based on an idea that there is a certain value to the status quo and that, when a plaintiff brings a legal claim, it disrupts the status quo, so the plaintiff should bear the risk of error when the evidence is in equipoise. See *Herman & MacLean v. Huddleston*, 459 U.S. 375, 390 (1983). But see CLERMONT, *supra* note 9, at 17–18 (dismissing this as no more justified than believing the *defendant* disrupts the status quo).

40. See 12 JOSEPH T. McLAUGHLIN & THOMAS D. ROWE, JR., MOORE’S FEDERAL PRACTICE § 60.43 (3d ed. 1997).

41. E.g., *Addington v. Texas*, 441 U.S. 418, 427, 431 (1979).

42. E.g., *Woodby v. INS*, 385 U.S. 276, 277 (1966).

43. E.g., *Schneiderman v. United States*, 320 U.S. 118, 123 (1943).

44. E.g., *Santosky v. Kramer*, 455 U.S. 745, 750, 769 (1982).

45. E.g., *Cruzan v. Mo. Dep’t of Health*, 497 U.S. 261, 265 (1990).

46. E.g., *N.Y. Times Co. v. Sullivan*, 376 U.S. 254, 279–80, 285–86 (1964).

preponderance of the evidence—we do not have a reason to favor errors for one party over the other.⁴⁷ Otherwise, the background assumptions in most civil trials include “that the litigants’ resources and abilities are approximately equal” and that in half the cases plaintiffs and defendants are each right about the disputed issues.⁴⁸ These assumptions, however, are fundamentally flawed.

Consider the following hypothetical cases. On the one hand, a breach of contract dispute between Amazon and FedEx. On the other hand, a janitor suing Wal-Mart for racial discrimination. Now ask yourself: In the janitor’s case, is it realistic to expect there to be an equality of resources and abilities between the parties? Assume further that Wal-Mart has been guilty for not paying overtime and other types of compensations and benefits before. Some might question whether the janitor case should be subject to the same error-distribution as the breach of contract case between two large corporations. This, however, is what occurs when we enforce the same standard of proof in both cases. These two examples illustrate situations in which the desirable error-distributions are likely to vary. This, in turn, suggests that the current evidentiary system fails to account properly for the relevant differences between different types of cases.

Cases might also differ for reasons other than economic disparity. Parties can have disparate knowledge of the relevant facts.⁴⁹ In a world with symmetrical information and low transaction costs, our assumptions about parties’ equality of resources and abilities might not carry much weight. However, in a world with informational and economic asymmetry and where discovery costs can be extraordinary, these assumptions become crucial. Ultimately, applying a rule that is based on assumptions that fail to hold in the real world can lead to unanticipated results. Instead of having errors equally distributed between parties, different litigants can end up carrying a larger share of the risk than what is socially desirable.

This problem is not unique to civil cases. In fact, in criminal law, the situation is even more extreme. There, not only are there widely diverse cases under only *one* standard of proof—beyond reasonable doubt—but also the Supreme Court has an elevated requirement to constitutional status. In *Winship*—and other cases addressing the Due Process Clauses of the Fifth and Fourteenth Amendments as a necessary element of a

47. See, e.g., *Addington v. Texas*, 441 U.S. 418, 427, 431 (1979).

48. Bell, *supra* note 19, at 570. But see generally Marc Galanter, *Why the “Haves” Come Out Ahead: Speculations on the Limits of Legal Change*, 9 LAW & SOC’Y REV. 95 (1974) (showing how “repeat players” can litigate cases and ultimately shape the development of law in more favorable fashions than “one shotters”).

49. Ronald J. Allen, *A Reconceptualization of Civil Trials*, 66 BOS. U.L. REV. 401, 427 (1986).

legally justified conviction in criminal adjudication—the Court made clear that the prosecution must establish every key element of their case beyond a reasonable doubt.⁵⁰ Justice Frankfurter went so far as to say that the standard of proof in criminal cases plays “a vital role in the American scheme of criminal procedure.”⁵¹

Behind the acquittal-friendly character of the beyond a reasonable doubt standard is the idea that mistaken convictions entail significantly higher costs than mistaken acquittals.⁵² The traditional view is that, because of this idea, the beyond a reasonable doubt standard needs to be set high across the entire criminal justice system.⁵³ Once we look closer, however, the idea that we value *all* mistaken convictions equally sounds normatively inadequate and descriptively inaccurate. Normatively, heavyweight considerations push us towards the desirability of enforcing different error distributions among substantially distinct types of cases.⁵⁴ Descriptively, social science research on jury decision-making suggests jurors already change decision-making thresholds in determining whether to condemn or acquit defendants based on a set of perceived relevant differences.⁵⁵

Consider a recently decided case where the Massachusetts Supreme Judicial Court reviewed which standard of proof the Massachusetts Sex

50. 397 U.S. 358, 363 (1970); *see also* Allen, *Winship*, *supra* note 11, at 39, 48–49. This principle actually goes back to the eighteenth century. *See* BARBARA J. SHAPIRO, *BEYOND REASONABLE DOUBT AND PROBABLE CAUSE: HISTORICAL PERSPECTIVES ON THE ANGL0-AMERICAN LAW OF EVIDENCE* 79 (1991); Bruce A. Antkowiak, *Judicial Nullification*, 38 *CREIGHTON L. REV.* 545, 560 (2005). The reasonable doubt standard applies across the board in the criminal justice system; it applies to misdemeanors and felonies, all degrees of offenses, and even in cases without a jury. *See* *United States v. Randolph*, 93 F.3d 656, 660 (9th Cir. 1996) (first citing *Jackson v. Virginia*, 443 U.S. 307, 319 (1979); and then citing *United States v. Mayberry*, 913 F.2d 719, 720 (9th Cir. 1990)).

51. *Winship*, 397 U.S. at 363 (“The reasonable-doubt standard plays a vital role in the American scheme of criminal procedure. It is a prime instrument for reducing the risk of convictions resting on factual error. The standard provides concrete substance for the presumption of innocence—that bedrock ‘axiomatic and elementary’ principle whose ‘enforcement lies at the foundation of the administration of our criminal law.’” (quoting *Coffin v. United States*, 156 U.S. 432, 453 (1895))).

52. *See* SHAPIRO, *supra* note 50, at 20–25; Erik Lillquist, *Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability*, 36 *U.C. DAVIS L. REV.* 85, 89 (2002); Anthony A. Morano, *A Reexamination of the Development of the Reasonable Doubt Rule*, 55 *B.U. L. REV.* 507, 509, 519 (1975).

53. This view is repeated by those on both sides of the legal spectrum. *See* Posner, *supra* note 38, at 411, 413; *see also* RICHARD A. EPSTEIN, *FORBIDDEN GROUNDS: THE CASE AGAINST EMPLOYMENT DISCRIMINATION LAWS* 225 (1992).

54. *See infra* Part IV.

55. *See infra* Part V.

Offender Registry Board (SORB) had to satisfy to classify a convicted sex offender under sex offender registry laws.⁵⁶ The plaintiff, a convicted sex offender, had been previously classified as having a moderate risk of re-offense by a preponderance of the evidence—the prevailing standard of proof according to state precedent at the time.⁵⁷ The court agreed with the plaintiff that recent amendments to the sex offender registry laws and other developments made the preponderance standard inadequate and concluded:

[D]ue process requires that a sex offender’s risk level be proved by clear and convincing evidence. The risk classification that SORB must make now has consequences for those who are classified that are far greater than was the case when we decided *Doe No. 972*. The preponderance standard no longer adequately protects against the possibility that those consequences might be visited upon individuals who do not pose the requisite degree of risk and dangerousness.⁵⁸

The important aspect of that case, for our purposes, is how the court reached its decision to increase the required standard of proof. Although the court noted it previously held that “the ‘possible injury to sex offenders from being erroneously overclassified’ was ‘nearly equal’ to ‘any harm to the State from an erroneous underclassification,’” it emphasized that significant revisions to sex offender registry laws and other developments imposed extra burdens on registered offenders without providing additional protections.⁵⁹ First, amendments had expanded the number of offenses that required registration.⁶⁰ According to the court, a higher number of

56. See *Doe v. Sex Offender Registry Bd. (Doe 380316 II)*, 41 N.E.3d 1058, 1071–72 (Mass. 2015). SORB is responsible for establishing and maintaining a “central computerized registry of all sex offenders required to register” pursuant to state legislation. MASS. GEN. LAWS ch. 6, § 178D (2018).

The file on each sex offender required to register [includes]:

- (a) the sex offender’s name, aliases used, date and place of birth, sex, race, height, weight, eye and hair color, social security number, home address, any secondary addresses and work address and, if the sex offender works at or attends an institution of higher learning, the name and address of the institution;
- (b) a photograph and set of fingerprints;
- (c) a description of the offense for which the sex offender was convicted or adjudicated, the city or town where the offense occurred, the date of conviction or adjudication and the sentence imposed;
- (d) any other information which may be useful in assessing the risk of the sex offender to reoffend; and
- (e) any other information which may be useful in identifying the sex offender.

Id.

57. *Doe 380316 II*, 41 N.E.3d at 1060.

58. *Id.* at 1061.

59. *Id.* at 1064 (quoting *Doe v. Sex Offender Registry Bd. (Doe 380316 I)*, 697 N.E.2d 512, 520 n.14 (1998)).

60. *Id.*

offenses increased the risk of errors in the SORB's task of classifying specific offenders' "risk of re-offense."⁶¹

Second, sex offenders faced increasingly rigorous reporting requirements and harsher sanctions for failing to meet such requirements.⁶² For instance, offenders were required to register secondary addresses, the names and addresses of the educational institutions they attend, and any new addresses ten days prior to moving as well were subjected to "intensive parole conditions."⁶³ If an offender failed to meet such requirements, he could be fined; however, if a judge determined imprisonment was a more appropriate sanction for failing to comply with registration requirements, the judge was bound to impose a mandatory minimum sentence of no less than six months.⁶⁴

Third, the court noted how offenders faced increasing difficulties based on their registered sex offender status.⁶⁵ These included limited ability to find work and housing, denial of access to federal social programs, as well as a vast range of other restrictions, such as a ban on "engaging in ice cream truck vending, regardless of whether their offense involved harm to a child" and a ban on living in a nursing home.⁶⁶

Fourth, the court focused on the fact that information about registered offenders was broadly disseminated.⁶⁷ For instance, both level two and level three sex offenders' information is posted online.⁶⁸ For the Justices, "[w]here previously the time and resource constraints of local police departments set functional limits on the dissemination of registry information, the Internet allows for around-the-clock, instantaneous, and worldwide access to that information—a virtual sword of Damocles."⁶⁹ This meant that if—after being classified as a level two or three offender and consequently having personal information posted online—an offender was later reclassified to level one such that the offender was no longer required to post his information online, the damage done to the offender's personal right would

61. *Id.*

62. *Id.* at 1065.

63. *Id.*

64. *Id.* at 1065–66 (citing MASS. GEN. LAWS ch. 6, § 178H (2010)).

65. *See id.* at 1070.

66. *Id.* at 1066.

67. *Id.* at 1064.

68. *See Moe v. Sex Offender Registry Bd.*, 6 N.E.3d 530, 538–39 (Mass. 2014).

69. *Doe 380316 II*, 41 N.E.3d at 1067 (citing *Moe*, 6 N.E.3d at 536–37).

be difficult to remediate.⁷⁰ This is particularly true because information posted online is hardly, if ever, truly withdrawn from the internet. Lastly, the court also raised a concern about whether the SORB’s guidelines for classifying offenders accurately reflected current scientific knowledge.⁷¹

In light of these factors, and after considering the effects of a new standard of proof in terms of more false negatives, the court determined the distribution of errors under the preponderance standard was tilted too sharply in favor of the government. For the court, such offenders “should not be asked to share equally with society the risk of error.”⁷² A higher standard of proof was necessary to preserve their due process rights and restore what the court saw as the proper balance of the distribution of risk.⁷³

Other cases might also benefit from higher standards of proof. In death penalty cases, the costs associated with wrongful conviction might be so high as to justify a standard of proof requiring “nearly absolute certainty.”⁷⁴ Where the repercussions from the conviction go beyond the formal sentence, 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 beyond a reasonable doubt.⁷⁵ Consider also cases involving low-level criminality, such as traffic offenses or other minor misdemeanors that would not result in any jail time; in such cases, the cost of an erroneous conviction seems lower compared to felony charges with long mandatory minimum sentences.

On the flipside, crimes where the costs from an erroneous acquittal are relatively high may require lower standards of proof. Consider a case involving a charge of terrorist acts where the defendant is a known public supporter of such practices and associated groups.⁷⁶ Here, the risk of harm

70. *Id.* (quoting Robert Kirk Walker, Note, *The Right to Be Forgotten*, 64 HASTINGS L.J. 257, 259–60 (2012)).

71. *See id.* (citing *Doe v. Sex Offender Registry Bd.* (*Doe 68549*), 18 N.E.3d 1081, 1092 (Mass. 2014)). This case is made even more interesting because it was decided at the same time as the federal government was requiring many universities around the country to lower the standard of proof in sexual harassment and sexual violence cases to *preponderance of the evidence* pursuant to Title IX, the federal law that prohibits gender discrimination in educational institutions that receive federal education funds. *See* CATHERINE E. LHAMON, OFFICE FOR CIVIL RIGHTS, U.S. DEP’T OF EDU., QUESTIONS AND ANSWERS ON TITLE IV AND SEXUAL VIOLENCE 5, 26 (2014) (“The school must use a preponderance-of-the-evidence . . . standard in any Title IX proceedings . . .”). The contradictory nature of both state-sanctioned recommendations in closely related cases is puzzling.

72. *Doe 380316 II*, 41 N.E.3d at 1071 (quoting *Addington v. Texas*, 441 U.S. 418, 427 (1979)).

73. *See id.* at 1072.

74. Lillquist, *supra* note 52, at 91.

75. *See id.*

76. For similar hypotheticals, see *id.*; Walen, *supra* note 14, at 421.

following from an erroneous acquittal seems higher than in traditional nonviolent drug offenses.⁷⁷ We can also imagine cases in which the benefits of an accurate *acquittal* might be low. For instance, there may be a case in which there is ample evidence that a particular defendant is engaged in drug trafficking, car theft, or stock manipulation, but there is relatively little evidence that the defendant committed the act for which the defendant is being charged. Society often accepts, or even demands, that more resources be spent on law enforcement to fight certain crimes, under the justification that fighting those crimes is more valuable than fighting others.⁷⁸ Likewise, society may decide the risks of adjudicative errors are more or less costly depending on the type of offense under consideration.

These claims might sound controversial. I suspect this is partially a result of the strong gravitational pull the current system exerts on our judgments. Let us consider a thought experiment: Imagine you are part of a special committee designing an evidentiary system from scratch. In particular, you are a member of a subcommittee charged with establishing the standards of proof applicable across different civil and criminal cases. You understand your task involves assessing the different significances and costs associated with false positives and false negatives—as well as benefits associated with accurate results—in different types of cases. It is implausible that such a committee would arrive at a unanimous decision that the proper ordering of costs and benefits is the same in virtually all cases. With that in mind, we can turn to our current evidentiary system and ask whether the current state of affairs with a fixed, very restricted number of standards applicable across drastically different cases is really justified.⁷⁹ This Article offers a different argument in support of answering

77. See Lillquist, *supra* note 52, at 162.

78. See, e.g., Talia Fisher, *The Boundaries of Plea Bargaining: Negotiating the Standard of Proof*, 97 J. CRIM. L. & CRIMINOLOGY 943, 953 (2007) (arguing for the desirability of a negotiable criminal standard of proof between prosecutors and defendants); Note, Winship on Rough Waters: *The Erosion of the Reasonable Doubt Standard*, 106 HARV. L. REV. 1093, 1095 (1993) (“[A]s society’s interest in crime control changes, society’s assessment of the proper balance between erroneous convictions and erroneous acquittals may change too.”).

79. Another interesting finding emerges from this quick thought experiment. It is not only difficult to justify a restricted number of standards of proof in civil and in criminal cases but it also becomes hard to justify a categorical distinction between the standards for civil and criminal cases. See, e.g., LAI, *supra* note 5, at 215–16 (“For all these reasons, it is difficult to justify a categorical difference in the standard of caution for civil and criminal cases; the standard in both contexts should be determined on the same broad principle. There must be as many ‘standards of proof’ as there are material differences in the circumstances

no to this question. We have good reasons to want different error distributions across varying circumstances.⁸⁰

I do not want to suggest society and the current legal system are completely oblivious to relevant differences between circumstances and cases. In fact, we constantly express our weighting of those differences through law. We assign higher penalties for what we believe are more serious crimes and lower penalties for less severe crimes. We hold defendants liable for higher damages—sometimes even exceeding plaintiffs’ losses—when their conduct is considered more blameworthy or socially costly. We have different regulations aimed at leveling the playing field between disparately situated members of society—labor law, consumer protection, antitrust. Even after a suit is filed, we express our different weighting of circumstances and parties through a myriad of evidentiary and procedural regulations—who carries the burdens of production and persuasion, exclusionary rules, rules of discovery, jurisdictional rules. To the extent these regulations make guilty or liable verdicts difficult to reach, they also have a potential impact on the distribution of errors in our society. Thus, we can see these regulations as mechanisms through which society already reflects the different weighing of the costs of errors in different situations, even if most legal rules are not justified concerning their potential impact on error-distribution.

An important question then follows: If society already reflects its views about the costs of errors in different situations through existing legal mechanisms, why should we discuss how to distribute errors with standards of proof? The answer is twofold. First, scholars and practitioners seldom acknowledge the impact of other legal rules in the distribution of errors. Second, even when we see discussions about standards of proof as a mechanism for distributing errors, we do not actually see discussions about how errors *should be* distributed. Instead, we uncritically accept that errors should be distributed similarly to vastly different situations as if we weigh the relative costs and significance of those situations similarly.

of cases—or, more accurately, there should only be one standard, a variant one. . . . Since criminal cases differ in both the gravity of charges and the range of punishment, there is no basis for applying to all of them a uniform standard [of proof].” (footnote omitted); Issachar Rosen-Zvi & Talia Fisher, *Overcoming Procedural Boundaries*, 94 VA. L. REV. 79, 86 (2008).

80. This proposal is not unprecedented in American legal scholarship. See, e.g., LAUDAN, *supra* note 11, at 55 (“Where standards of proof are concerned, I am not convinced that one size fits all.”); Kaplow, *Burden of Proof*, *supra* note 4, at 786 n.86 (“The . . . optimal evidence thresholds and the evidence threshold . . . will vary greatly by context, even at fairly refined levels.”); Walen, *supra* note 14, at 417 (“[W]hether one is a consequentialist or deontologist, one has good reason not to cling artificially to the idea that only one [standard of proof] must be used for all of criminal law.”).

But once we consider other legal rules, we acknowledge how different those situations are and enact a long list of regulations that reflect our sensitivity toward those differences. At that point, however, discussions about the impact of error distributions are nowhere to be found. We, therefore, end up never really discussing, or even considering, the impact of legal rules on the distribution of errors in our society and what the socially desirable distributions actually are. We need to bring such discussions to the surface. And what better time to have these than when thinking about the very mechanism that many scholars and practitioners alike agree has error-distribution as a main function?

IV. NORMATIVE ARGUMENTS IN FAVOR

This Part surveys normative arguments in favor of varying standards of proof. It starts by describing recent influential works on standards of proof by consequentialist legal scholars. It then argues that once we follow the logic of their positions to its ultimate conclusions, we must consider seriously different standards for different types of cases. It then discusses how the same is true for arguments based on fairness considerations. These share similar conclusions with welfare-based consequentialist arguments, even though they begin with different premises. Finally, this Part explores how arguments premised on considerations regarding the division of legal entitlements in our society also support moving to a system with varying standards of proof.

A. Welfare Considerations

The topic of standards of proof has recently regained currency as a subject among economically-minded legal scholars.⁸¹ One of their focuses has

81. See generally, e.g., Michael L. Davis, *The Value of Truth and the Optimal Standard of Proof in Legal Disputes*, 10 J.L. ECON. & ORG. 343 (1994); Dominique Demougin & Claude Fluet, *Deterrence Versus Judicial Error: A Comparative View of Standards of Proof*, 161 J. INSTITUTIONAL & THEORETICAL ECON. 193 (2005); Edward K. Cheng, *Reconceptualizing the Burden of Proof*, 122 YALE L.J. 1254 (2013); Kaplow, *Burden of Proof*, *supra* note 4; Louis Kaplow, *Information and the Aim of Adjudication: Truth or Consequences?*, 67 STAN. L. REV. 1303 (2015); Louis Kaplow, *Optimal Proof Burdens, Deterrence, and the Chilling of Desirable Behavior*, 101 AM. ECON. REV. 277 (2011); Thomas J. Miceli, *Optimal Prosecution of Defendants Whose Guilt Is Uncertain*, 6 J.L. ECON. & ORG. 189 (1990); A. Mitchell Polinsky & Steven Shavell, *Legal Error, Litigation, and the Incentive to Obey the Law*, 5 J.L. ECON. & ORG. 99 (1989); Daniel L. Rubinfeld & David E.M. Sappington, *Efficient Awards and Standards of Proof in Judicial*

been on how to set the levels of standards to maximize social welfare.⁸² More specifically, these scholars have presented thought-provoking analyses of how the optimal level of standards depends on a range of elements that are bound to vary in different situations. My goal is to show that, regardless of specific criticisms or departures from more traditional accounts, many of the key elements of these projects support a system of varying standards of proof.

Louis Kaplow has delivered what is perhaps the most comprehensive recent welfare analysis of standards of proof.⁸³ For that reason, this section draws heavily from his framework. Kaplow's analysis is based on the idea that the most relevant effects of setting the standard of proof at different levels are the changes in the costs associated with "deterrence and chilling" in society at large.⁸⁴ Although deterrence and chilling refer ultimately to the same phenomenon—the impact of legal rules on primary behavior—the use of different terms is meant to capture the distinction between what Kaplow refers to as "harmful acts"—acts with negative net social benefits—and "benign acts"—acts with positive or zero net social benefits.⁸⁵

The essential premise is that the prospect of sanctions will influence the behavior of individuals thinking about committing harmful or benign acts.⁸⁶ More precisely, individuals with an opportunity to commit a harmful or benign act are disincentivized when the private benefit an individual expects to gain from committing the act is *lower* than the sanction an

Proceedings, 18 RAND J. ECON. 308 (1987). For earlier examples, see generally Alan D. Cullison, *Probability Analysis of Judicial Fact-Finding: A Preliminary Outline of the Subjective Approach*, 1 U. TOL. L. REV. 538 (1969); John Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065 (1968); Posner, *supra* note 38. This type of work is not without its critics. See, e.g., Walen, *supra* note 14, at 423 (arguing consequentialist approaches to standards failed to "give the distinction between innocence and guilt its proper moral significance").

82. See generally Kaplow, *Burden of Proof*, *supra* note 4. It is important to note that Kaplow's account differs drastically from the formulation of standards in Part II. For him, the account focused on the distribution of errors between parties suffers from the defect of not taking into consideration all the elements included in his welfare analysis. See generally *id.* It is out of the scope of this article to arbitrate between different formulations of standards of proof. My objective in Part II was to introduce the reader to a widespread way of formulating standards of proof according to scholars and practitioners alike. My goal with introducing Kaplow's welfare analysis in this Part IV is to highlight how it also leads to the conclusion that a system with varying standards of proof is normatively desirable. I take no stance on which formulation is preferable.

83. See generally *id.*

84. *Id.* at 746, 746 n.16.

85. *Id.* at 747.

86. See generally *id.*

individual expects to receive for committing the act.⁸⁷ That is, even in deciding whether to commit an act that is currently lawful, individuals weigh the expected private benefit against the expected private costs of committing that act.⁸⁸

Ideally, only harmful acts would be deterred, while benign acts would not be chilled. The fact that some benign acts can be confused with harmful acts makes this result difficult. One important question for our purposes is what happens to the levels of deterrence and the chilling effect of different acts when we change the applicable standard of proof? Asking this question will allow us to understand better some of the costs and benefits involved in the choice of standards. Recall the discussion above about the trade-off involved when we alter the level of the standards of proof.⁸⁹ We saw that if we lower the standard, that is, move the line in Figures 2 or 3 to the left, we increase the likelihood that the legal system will correctly assign liability to truly guilty individuals that go to trial while at the same time increasing the likelihood that truly innocent individuals will mistakenly be sanctioned.⁹⁰

Likewise, when we lower the standard, holding other features of the legal system constant, we increase the chances of punishing both harmful *and* benign acts, which, in turn, increases the expected sanction for both types of acts.⁹¹ As a result, harmful acts will be deterred while more benign acts will be chilled.⁹² When harmful acts are deterred, society benefits; when benign acts are chilled, society loses. Thus, we must compare the aggregate deterrence benefits to the overall chilling costs.⁹³ The important question from a welfare-based analysis is whether the net benefit from these effects is positive or negative. If the benefit resulting from the deterrence of harmful acts is greater than the loss resulting from the chilling of benign acts, then reducing the level of standard is desirable.⁹⁴ Whereas if the reverse is true—that is, if the loss outweighs the benefits—not only would such reduction be socially detrimental but also an increase in the

87. *See generally id.* This is the same regardless of whether the sanction is correctly or mistakenly imposed on the person. *Id.* at 749–50.

88. *Id.*

89. *See supra* pp. 171–72.

90. *See supra* pp. 171–72.

91. Kaplow, *Burden of Proof*, *supra* note 4, at 763.

92. *Id.*

93. *Id.*

94. *Id.*

level of the standards would be “socially desirable.”⁹⁵ The optimal standard of proof for welfare maximization is given by the point where these two opposing forces reach an equilibrium.⁹⁶

To appreciate fully the point that the optimal standard varies according to the act under consideration, we must break down the two forces explained in the last paragraph into its defining components. Let us start with the deterrence effect. The level of deterrence following a change in the level of the standard depends primarily on three factors: “the increase in expected sanctions, . . . the concentration of marginal harmful acts,” and the difference between social costs and the private benefits of marginal harmful acts.⁹⁷

First, the change in expected sanctions. This is a product of three other factors: the (1) likelihood that the competent authorities will identify the harmful act, (2) likelihood that the individual will be subject to sanctions, and (3) magnitude of the sanctions imposed.⁹⁸ A change in the level of the applicable standard directly impacts the second factor—the likelihood that the individual will be subject to sanctions.⁹⁹ The exact impact that a change in the level of the standard has on this factor can be determined from Figures 2 or 3. Depending on whether we increase or lower the standard—move the line to the right or move it to the left, respectively—the likelihood that the individual will be subject to sanctions will be equal to the area under the harmful acts curve.

We can then calculate the impact of a change in the level of the standard by determining the difference between the new area under the curve after we changed the standard and the area under the curve before we moved the line. For example, as one raises the level of the standard, we move the point representing the standard to the right in Figures 2 or 3. As a result, parts of the area under the harmful acts curve that used to be to the right of the standard line will now be to the left. That the new area under the harmful acts curve to the right of the standard line is now *smaller* than the area under the curve to the right of where the line used to be means that the likelihood that the individual will be subject to sanctions is *lower*. We can then compute the increase in the expected sanction by multiplying the magnitude of the sanctions imposed by the likelihood that (1) the competent authorities will identify the harmful act and (2) the individual will be subject to sanctions. Assuming both likelihoods stay constant, an *increase* in the

95. *Id.*

96. *Id.*

97. *Id.* at 764.

98. *Id.* at 754.

99. *Id.*

level of the standard generally means the expected sanction for harmful acts is *lower*.

The second component that will help determine the deterrence benefits is the “concentration of marginal harmful acts.”¹⁰⁰ This is determined by “how often private benefits from [those] acts fall [within the] range” of the difference of the increase in the expected sanction.¹⁰¹ Together with the first component, this provides the change in the number of individuals who will be deterred because of a legal change, such as adjusting the level of the applicable standard of proof.¹⁰² To reach that figure, we must consider which specific acts will—or will not—be deterred as we alter the standard.¹⁰³ The acts that will be deterred “will be those that generate private benefits” lower than the new expected sanction after lowering the standard, minus those acts which were already deterred under the old expected sanction.¹⁰⁴

The third and last factor determining deterrence effects of changes in the standard is “the net social gain per deterred act.”¹⁰⁵ This is equal to the “difference between social harm . . . and the private benefit” of a marginal harmful act.¹⁰⁶ The size of the social harm will depend on the context: “for littering, it will be small, but for discharging highly toxic chemicals into the water supply, it will be large.”¹⁰⁷ And, the benefit of a marginal harmful act “equals the expected sanction.”¹⁰⁸ “From the social harm per act, we [still] need to subtract each act’s private benefit.”¹⁰⁹

The exercise is not over yet. “Against th[e] deterrence benefit[s], we must weigh the chilling cost[s].”¹¹⁰ These costs are equal to “the increase in the expected sanction[s] for benign acts weighted by the concentration of marginal benign acts.”¹¹¹ This “product is then multiplied by the net cost per act that is chilled.”¹¹² The analysis is analogous to the deterrence

100. *Id.* at 764.

101. *Id.* at 765.

102. *Id.* at 765–66.

103. *Id.* at 765–67.

104. *Id.* at 765.

105. *Id.* at 766.

106. *Id.*

107. *Id.*

108. *Id.*

109. *Id.*

110. *Id.* at 768.

111. *Id.*

112. *Id.*

effects discussion above, so I will leave to the reader the task of filling in the blanks.

The optimal standard will be where the marginal “deterrence benefit just equals the chilling costs.”¹¹³ At that equilibrium, society does not have an incentive to increase or lower the applicable standard. Certainly, this calculation in the real world is highly complex. Much depends on many empirical facts that are likely to vary quite drastically in context. The optimal level of the standard is determined by a multitude of factors, which can take on a vast range of values. The important takeaway for this Article is that we have reason to believe deterrence and chilling costs will be vastly different among different types of acts.¹¹⁴ The conclusion that follows is that the optimal level of the standard also varies according to different acts. As Kaplow puts it:

In some cases, the optimal evidence threshold might be quite high. For example, if the act is not very harmful, there are many benign acts that look like the harmful act, and the legal system does not discriminate well between them, then chilling costs will exceed deterrence benefits except possibly at very high evidence thresholds. On the other hand, if the act is quite harmful, few benign acts are likely to be confused with harmful acts, and it is otherwise difficult to generate a high expected sanction (perhaps stage one only identifies a small fraction of harmful acts), then deterrence benefits will exceed chilling costs until the evidence threshold is reduced substantially.¹¹⁵

The fact that the optimal standard varies for different acts gives us a reason to question whether the current system of evidentiary rules, in which different acts are lumped together under the same standard, is really the best system for welfare maximization analyses. It does not take much institutional imagination to envision a system that tracks the optimal level of the standard more closely than our current system does.

This framework is not without its problems. For instance, it potentially requires high administrative costs and ignores effects of dynamic behavior.¹¹⁶ I deal with this and other objections below. Here, it suffices to say that I do not endorse every aspect of Kaplow’s account of standards of proof. My objective in exploring this account was to illustrate one type of reasoning about evidence law—in general—and standards of proof—in particular—that stresses the sometimes drastically different private and social costs associated with different types of conduct and how these costs should weigh in our thinking about how to regulate behavior.

113. *Id.* at 769.

114. It is also important to keep in mind that other enforcement mechanisms have behavioral effects, and the interaction of changes in part of the system might have unpredicted consequences—think second best-type arguments.

115. Kaplow, *Burden of Proof*, *supra* note 4, at 770.

116. Allen & Stein, *supra* note 9, at 580, 582–84.

B. Fairness Considerations

In his only published work on procedural and evidentiary issues, Ronald Dworkin explores whether a community that provides its citizens a right not to be convicted if innocent but also decides questions of procedure and evidence according to an ordinary welfare analysis, can be blamed for inconsistency.¹¹⁷ He starts by addressing the following question: If we recognize “that people have a right not to be convicted of . . . crime[s]” they did not commit, does it follow “that people have a right to the most accurate procedures possible” to prove their guilt or innocence, regardless of how expensive these might be?¹¹⁸

Suppose cases could be made marginally more accurate if we had a “best out of three” rule. Under this rule, cases would be tried by up to three different juries and the party that got favorable verdicts from two juries would prevail.¹¹⁹ Suppose further that this would cost cases more time and money. If society decides to continue to have only one trial to save costs, it will forgo the marginal increase in accuracy. Consequently, courts will mistakenly convict some innocent defendants that otherwise would be acquitted under the best out of three rule. Is the decision not to adopt such a rule unjust against defendants who are denied a reduction in their risk of being mistakenly convicted? If yes, we would be forced to admit our justice system is largely unfair because we do not implement all possible instruments that could increase, however marginal, accuracy regardless of the instruments’ costs.

Most people would resist conceding this last point. Should we, then, hold that people at trial have no rights to any particular level of accuracy? That seems to be our assumption if we left all procedural questions to cost-benefit analysis, like the one in the last section. But would that position be consistent with the initial assumption that people have a right not to be convicted if innocent? Can we find some middle ground between the

117. See DWORKIN, *supra* note 5, at 72–80.

118. *Id.* at 79–80. Note that a right in a Dworkinian sense is quite strong, implying that arguments of policy are not sufficient to prevent the enforcement of a right. See *id.* at 72. For instance, “[i]f a prosecutor were to pursue the person he knew to be innocent, it would be no justification or defense that convicting that person would spare the community some expense or in some other way improve general welfare.” *Id.* When a person asks a court to enforce a right, “that the community would be better off if that right were not enforced, is not [to be] counted a good argument against him.” *Id.* at 73.

119. Or we could add the counts on individual cases of actions. The details are not important for this quick thought experiment.

absurd claim that people have a right to all procedures that increase accuracy regardless of costs and the seemingly nihilistic claim that people lack a right to any level of accuracy?¹²⁰

Dworkin attempts to answer these difficult questions through the concept of “moral harm.”¹²¹ According to him, we must distinguish between (1) the harm a person suffers through punishment—pain, frustration, damages to reputation, loss of income—that we might call “bare harms,” and (2) another kind of harm a person might suffer whenever the punishment is unjust or otherwise treated unjustly—whether the person knows, or cares, about it.¹²² It is this second type of harm Dworkin refers as “moral harm.”¹²³ We can grasp that idea when we sympathize for someone upon learning that the person was treated unjustly or cheated, even though we learned nothing more about how much that person suffered from punishment.¹²⁴

With the notion of moral harm, we can better understand the supposed strangeness behind the behavior of a society that recognizes that people have a right not to be convicted if innocent but decides procedural and evidentiary policies according to welfare analyses. It seems inconsistent to acknowledge the right not to be convicted if innocent unless we recognize moral harm as a special kind of harm against which people must be protected. The problem is that typical welfare analyses seem to have no place for moral harm.¹²⁵

Importantly, even an analysis that *has* a place for moral harm should not strive to reduce it to zero. “[W]e do not lead our lives to achieve the minimum of moral harm at any cost”¹²⁶ Rather, we accept weighty

120. Note, it is not true that the right not to be convicted if innocent is meaningless in a society that does not recognize a right to any particular level of accuracy and decides its procedural rules solely on the basis of cost-benefit analysis. That right still protects citizens against unfounded suits.

121. DWORKIN, *supra* note 5, at 80–90.

122. *Id.* at 81.

123. *Id.* Someone might question this distinction by saying it confuses the quantity of harm someone suffers from a punishment with the different issue of whether that harm is just. When punishment is unjust, the harm suffered is unjust, but it is confusing to say the injustice in some way adds to the harm. However, one need not accept this distinction to acknowledge the substantive point behind the idea of moral harm—that someone suffers a special injury when treated unjustly. *See id.* at 80–81.

124. *See id.* Although Dworkin does not discuss this possibility, it seems plausible that the concept of moral harm could be modified slightly to cover cases of wrongful acquittals.

125. Dworkin here is referring to “utilitarian calculations” that measure costs and benefits according to a utilities and disabilities function determined largely by some psychological states along a pleasure and pain spectrum, even if such function includes people’s preferences to be punished, or have others be punished, unjustly. *Id.* at 82. This is because, to Dworkin, the idea of moral harm is an objective notion. *See id.* at 80–86.

126. *Id.* at 86.

“risks of suffering injustice[s]” to achieve other worthy goals.¹²⁷ We do so when we enter into relationships, sign contracts, or trust friends.¹²⁸ Likewise, society should have a similar approach to moral harm.¹²⁹ Under this light, society’s evidentiary rules are embedded in its choices about the relative importance of different types of moral harm.¹³⁰

Now we get to the heart of why I believe this discussion provides an argument for a system with varying standards of proof. Nothing in Dworkin’s analysis forces us to place the same importance on all types of moral harms in different cases. Just like we expect that, in a fair society, the level of sanctions attached to various crimes should be consistent with the relative importance society places on that crime, we should also expect that procedural and evidentiary instruments correlate to the relative importance of risk of moral harm. If we value the injustices that would follow from wrongful convictions or acquittals differently, then we should weigh moral harm differently among different cases. According to this framework, we should have different standards of proof for different situations, and these standards should reflect society’s choices about the relative importance of the types of moral harm involved.

There are many different instruments a society might use to alter its legal procedures to have them reflect the relative weight of moral harms. One way for society to pay a higher price for the accuracy of its adjudicative proceedings is to guard against particular kinds of mistakes that involve greater moral harm. Standards of proof are the clearest instrument through which this protection can be articulated and achieved. In Dworkin’s words:

[I]n the civil law . . . it is generally assumed that a mistake in either direction involves equal moral harm. But [the different standard in defamation suits] may represent some collective determination that it is a greater moral harm to suffer an uncompensated and false libel than to be held in damages for a libel that is in fact true.¹³¹

In a Dworkinian universe, these considerations acquire additional strength because people might have a right to procedures that correlate the *correct*

127. *Id.*

128. *Id.*

129. *See id.*

130. *Id.* (“[W]e might regard the design of criminal and civil procedures as a fabric woven from the community’s convictions about the relative weight of different forms of moral harms, compared with each other, and against ordinary sacrifices and injuries.”).

131. *Id.* at 89.

importance to the risk of moral harm they face.¹³² This is because, for Dworkin, it is part of the principle that any political decision must treat all citizens as equals, and no decision may deliberately impose a greater risk of moral harm on any citizen than that which it imposes on any other.¹³³ The difficulty, however, is that all members of society do not face equal risks of suffering the same kinds of moral harms. People are not equally likely to be drawn into the criminal process though innocent or equally likely to benefit from the savings gained by choosing a particular rule of evidence rather than a socially more expensive rule. Economic inequality and existing prejudices are factors that might make it likely that members of different segments of society will be more likely to be accused of certain types of crimes. For example, the poor might be more likely to be accused of petty thefts and the rich of antitrust violations. Other differences, such as in temperament and personalities, also make it more likely that certain types of people will be accused of different crimes. The hot-blooded or the gun aficionados are more likely to be accused of violent or gun-related offenses, while the greedy or Wall Street-types are more likely to be accused of securities violations.

Because of these inequalities, the final weighing of moral harms against bare harms is likely to be controversial. Minorities, who might face a higher probability of being drawn into the criminal process even though innocent, will protest that the level of accuracy provided by the rules at play is too low, undervaluing the moral harms following from unjust convictions. The majority, on the other hand, might think the level of accuracy is too high, and thus overvalues moral harms compared to the benefits forgone when using taxpayer funds for other purposes.

Overall, this means procedural and evidentiary rules involve important political decisions. They involve decisions about weighing of moral harms against bare harms, which crucially affects the probabilities of different people and social groups. These decisions also involve issues about the use of taxpayers' money. It is expected that people will gain differently from a different use of public funds. In turn, this suggests it is not enough to fault a society's procedural and evidentiary system because it imposes different consequences on different members. Society's decisions about its procedural and evidentiary rules can only be accused of being *unfair* if these decisions constantly discriminate against independently distinct

132. *Id.* at 89–90. To say this is a right is to say that once the content of the right is determined—that is, once the appropriate evaluation of the moral harm is done—society must protect the right of its citizens to procedures that represent that evaluation, even if the general welfare suffers as a consequence. *See id.*

133. *Id.* at 85.

groups.¹³⁴ It is not enough to consider these decisions unfair simply because they weigh moral harms differently from what other members of the society would weigh them.

Dworkin also mentions that people have a right to a “consistent” weighing of the importance of moral harm.¹³⁵ This consistency is not merely that *like cases should be treated alike*. There also should be consistency with the community’s evaluation of the moral harm at play.¹³⁶ For Dworkin, to assure consistency, we must search society’s textual and historical record to find an interpretation that fits and justifies most—but not necessarily all—of that record.¹³⁷ Most of the time, this consistency requirement might act as a conservative force protecting the accused from changes in the evaluation of moral harm. But it can also act as an instrument for reform by identifying existing, and even deeply embedded, procedures as mistakes that cannot be made consistent with legal and political practices.

How much consistency is required? Should we require the same weight of moral harm across all cases? I do not think so. Consistency across the board would require that society value different moral harms as having the same weight. Yet this is not what we observe. The different use of standards of proof between civil and criminal cases suggest that society assigns disparate values to moral harms following from different cases. However, even within criminal or civil cases, we find evidence for various evaluations of moral harms through procedural and evidentiary mechanisms that make conviction more or less likely as well as increase or decrease the interval between the minimum and maximum sanction. Examples include evidentiary privileges, class actions, and mandatory minimums. It must then be the case that the consistency requirement mentioned above can be fulfilled even if procedures and evidentiary rules vary among different types of cases. This is enough for the purpose of the proposal in the Article.

One potential objection arises out of fairness considerations. To some, it might seem that fairness considerations give rise to equality concerns. Equality concerns suggest that “anyone who brings a civil claim should

134. *See id.* at 87–88.

135. *Id.* at 90.

136. *See id.* Both of these rights are rights in the strong Dworkinian sense, according to which rights act as a trump over the balance of costs and benefits. *See id.*

137. *Id.* at 96 (“The basic procedural right in civil litigation is the right that the risk of the moral harm of an unjust result be assessed consistently so that no less importance is attached to that risk by a court’s procedural decision that is attached in the law as a whole.”).

have it judged by the same standard of proof that applies to everyone else[.]”¹³⁸ But this objection confuses equal treatment with being treated equally. We ought to differentiate between these two ideas. To be treated equally is not equivalent to receiving the same number of goods or the same level of burdens as everyone else.¹³⁹ Rather, it is better understood as being treated with the same degree of concern and respect.¹⁴⁰ Being treated equally does not demand a single fixed standard of proof to govern vastly different cases. In fact, being treated equally seems to require the exact opposite. It requires that individuals have a right to procedures that correlate the *correct* importance to the risk of moral harm they face. In that sense, a system with varying standards of proof seems to be a dictate of equality.

It is unsurprising that Dworkin’s analysis is sharply distinct from Kaplow’s. The point of this Article is not to offer one univocal defense of the proposal for varying standards of proof. Rather, the point is to discuss different types of arguments for this proposal.¹⁴¹

C. Distributive Considerations

Distributive analyses are almost always highly controversial. Perhaps for that reason, distributive considerations are often disguised as well-known considerations, such as formal legal equality or equality of bargaining power.

138. LAI, *supra* note 5, at 222 (citing C.M.A. McCauliff, *Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?*, 35 VAND. L. REV. 1293, 1334–35 (1982)); *see also* Walen, *supra* note 14, at 426–27 (arguing (1) against instrumental considerations in setting standards of proof and (2) that retributive considerations should lead us to uphold the same beyond a reasonable doubt standard across much of criminal law). *But see* Jeffrey Reiman & Ernest van den Haag, *On the Common Saying that It Is Better that Ten Guilty Persons Escape than That One Innocent Suffer: Pro and Con*, SOC. PHIL. & POL’Y, Spring 1990, at 226, 227–28, 247–48.

139. *See* RONALD DWORKIN, *TAKING RIGHTS SERIOUSLY* 227 (1978).

140. *Id.*

141. For another interesting analysis that could be included under this section—and that also shows some skepticism towards a system with fixed standards—see LAI, *supra* note 5, at 215–23 (“For all these reasons, it is difficult to justify a categorical difference in the standard of [proof] for civil and criminal cases; the standard in both contexts should be determined on the same broad principle. There must be as many ‘standards of proof’ as material differences in the circumstances of cases—or, more accurately, there should only be one standard, a variant one.” (footnote omitted) (citing G.H.L. Fridman, *Standards of Proof*, 30 Canadian Bar Rev. 665, 670 (1955))). “Since criminal cases differ in both the gravity of charges and the range of punishment, there is no basis for applying to all of them a uniform standard of [proof].” *Id.* at 216; *see also* David Hamer, *Review: A Philosophy of Evidence Law—Justice in the Search for Truth*, 13 INT’L J. EVIDENCE & PROOF 161, 161 (2009) (book review).

Rarely does someone explicitly advocate for a legal change because it transfers resources away from one social group to another.¹⁴²

The kind of distributive analysis I am interested in here focuses on the role of background entitlements in the distribution of resources among different social groups.¹⁴³ The division of entitlements in society matters greatly for the expected consequences of legal changes. Any argument for the welfare virtues of any legal change necessarily depends on specific assumptions about the background entitlements of those affected.¹⁴⁴ This is “[b]ecause entitlements are a component of wealth.”¹⁴⁵ Entitlement settings “influence the allocation of resources through ‘wealth effects’”—that is, through composition of demand, price, and income elasticity of demand.¹⁴⁶ So, if we can manipulate background entitlements, we can bring about changes in wealth so great that they should allow us to influence greatly—as to dominate potentially—the cost-benefit analysis of any legal change.¹⁴⁷

Most dramatically, it might turn out that the setting of the entitlement completely determines *ex ante* the outcome of the welfare analysis.¹⁴⁸ For example, in situations in which there can be no voluntary transactions between the individuals affected by a proposed legal change—because of prohibitive transaction costs, perhaps—the initial entitlement allocation will dictate the analysis outcome. If firms are entitled to pollute, people that own land nearby and are negatively affected by the firms’ pollution might not be able to transact with the polluters to reduce the firms’ externalities

142. Kennedy, *Distributive and Paternalist*, *supra* note 7, at 588. Duncan Kennedy makes this point eloquently:

Peace and happiness seem to require that most of the time we not think at all about the justice of distributive shares. Otherwise, we risk falling into depression, or into a rage, and, in either case, out of sympathy with one another. Or worse yet, we may fall into the kind of sympathy—intense but selective—that leads to civil war. There is therefore a taboo on the explicit consideration of distributive consequences, let alone distributive goals.

Id.

143. For a seminal treatment of entitlements and disablements and their correlatives, see generally Wesley Newcomb Hohfeld, *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 26 *YALE L.J.* 710 (1917); Wesley Newcomb Hohfeld, *Some Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 23 *YALE L.J.* 16 (1913).

144. See Kennedy, *Cost-Benefit Analysis*, *supra* note 7, at 388.

145. *Id.* at 423.

146. *Id.*

147. *Id.* at 423–26.

148. *Id.* at 426.

because of high transaction costs. Consequently, a rule that allows firms to pollute would be seen as effecting an *efficient* allocation of resources.¹⁴⁹ Inversely, if landowners were entitled to clean air, firms would have to transact with landowners to engage in pollution-causing activities. But if transactions costs are prohibitively high, these transactions will not occur. As a result, a legal rule that prohibits firms from polluting would seem *efficient*.

Entitlements can have even more subtle effects. If landowners are entitled to clean air, that makes their position more valuable—for instance, in terms of higher real estate value. This, in turn, increases the amount landowners are willing accept to allow pollution from nearby firms. Conversely, if firms are entitled to pollute, they are better-off and, consequently might start to value this entitlement at a higher selling price. If this price is greater than landowners' willingness to pay, there would be no gains from trade. Thus, the initial allocation of entitlements will not only remain intact but will also appear to be the efficient allocation under a welfare analysis. Theoretically, there will always be a particular combination of background entitlements that generates wealth effects so great as to render any change inefficient. This introduces a strong bias in favor of the *status quo* in this type of analysis. Moreover, because we often have no good reason to set the initial entitlement allocation for one party over another, this might be seen as introducing arbitrariness to the analysis.

Distributive analyses are at the fringes of legal scholarship.¹⁵⁰ Evidence and procedural laws are often left untouched. I believe this is a mistake. Just like private law, evidentiary and procedural mechanisms should also constitute individuals' background entitlements.¹⁵¹ The exact impact of changing the applicable standard on buyers and sellers will partially depend on the price and income elasticities of supply and demand curves as well as the competitive structure of that specific market. There might also be

149. See generally R.H. Coase, *The Problem of Social Cost*, J.L. & ECON., Oct. 1960, at 1 (1960).

150. Noteworthy exceptions include: RONALD DWORKIN, SOVEREIGN VIRTUE: THE THEORY AND PRACTICE OF EQUALITY 307–19 (2000) (health-care policy); LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE (2002) (providing an in-depth discussion of the role of distributive considerations in the assessment of legal policies); Matthew D. Adler & Eric A. Posner, *Rethinking Cost-Benefit Analysis*, 109 YALE L.J. 165 (1999) (administrative practice); Carlos A. Ball, *Autonomy, Justice, and Disability*, 47 UCLA L. REV. 599 (2000) (disabilities); Daniel Markovits, *How Much Redistribution Should There Be?*, 112 YALE L.J. 2291 (2003) (developing a new account of egalitarianism based on a new conception of nonsubordination); and Edward J. McCaffery, *The Uneasy Case for Wealth Transfer Taxation*, 104 YALE L.J. 283 (1994) (considering distribution implications for the choice of the tax base).

151. See generally Symposium, *Presumptions and Burdens of Proof*, 17 HARV. J.L. & PUB. POL'Y 613 (1994).

distributive effects on third parties, who can be made better off as a consequence of lower standards because the increased expected sanctions incentivize firms to reduce the risks of their economic activity, which, in turn, might involve negative externalities.

The conceptual point here is that, similarly to rules of private law, we can manipulate evidence rules operating in the background to bring about changes in wealth with noticeable effects for a welfare analysis. We can also manipulate evidence rules that bring about changes in wealth so extreme that the change will determine the outcome of the analysis from the outset. Consider what would happen if potential buyers are so impoverished that they cannot pay a higher price for the product once the industry faces higher expected accident costs due to a reduction in the standard of proof for consumer product mass tort cases. Or consider the consequences for private bargaining of a system with prohibitive judicial costs.

The point of this section is to argue that this distributive analysis of evidence law also supports a system with varying standards of proof. If the level of the applicable standard is seen—together with the rest of the constellation of evidentiary and procedural regulation—as part of individuals' background entitlements with relevant consequences for the distribution of wealth and income in our society, then any legal reformer intending to alter that distribution of wealth or income among different social groups can also make use of standards of proof or evidentiary and procedural regulation to reach his objectives. This is true regardless of whether distributive considerations are instrumental to attaining other goals. It is possible, albeit politically controversial, that one might attempt to promote a specific type of legal reform with the sole objective of altering the distribution of wealth or income between social groups. For instance, the example above involving changing standards in mass consumer tort cases could be part of a greater political strategy to reallocate funds from certain groups to others. Moreover, even if such a proposed legal change is not publicly defended on such grounds, the wealth effects following it constitute predictable consequences which should be acknowledged and included in the public debate.

This might surprise some. Evidence law has traditionally been perceived as a dry and highly technical field. Not surprisingly, it was thought to belong almost exclusively to the professional interest of practitioners. But nothing could be further from the truth—an especially problematic situation for a field that has seeking truth as one of its main goals.

V. A POSITIVE ARGUMENT IN FAVOR

As I mentioned in Part II, we believe we are indifferent between errors for plaintiffs or defendants in most civil suits and that we accept the same error distribution for all criminal cases. The previous section presented reasons why, in many circumstances, we might not want to do what we say. This section argues that what we do is different from what we say. Our practice is out of tune with our discourse. I attempt to show this by referencing social science research on jury decision-making.

Although only a small percentage of cases are affected by jurors' decisions,¹⁵² the jury has received more attention from behavioral scientists and other researchers than any other comparable decision-making institution.¹⁵³ Although there were early twentieth century anecdotal studies, "systematic and [rigorous] empirical inquiry into . . . jury decision-making [can be] traced to [the 1950s] University of Chicago Law School's Jury Project."¹⁵⁴ That project hosted a series of investigations into the role and functions of juries in the American legal system, gathering information from over 3,500 criminal jury trials.¹⁵⁵ Among other things, researchers compared the verdicts rendered by juries with hypothetical verdicts that judges presiding the juries would have rendered, revealing that juries and judges agreed roughly 75% of the time.¹⁵⁶ Most disagreements stemmed from juries' greater lenience in comparison to judges.¹⁵⁷

Studies on jury decision-making increased in the 1970s because of, in part, critical decisions from the U.S. Supreme Court.¹⁵⁸ In *Witherspoon*

152. In 2017, 327,557 cases were filed and 347,713 cases were terminated in all U.S. District Courts. U.S. COURTS, JUDICIAL FACTS AND FIGURES, U.S. DISTRICT COURTS—TOTAL CIVIL AND CRIMINAL CASES FILED, TERMINATED, AND PENDING Table 6.1 (Sept. 30, 2017), https://www.uscourts.gov/sites/default/files/data_tables/jff_6.1_0930.2017.pdf [<https://perma.cc/5AB6-2F2G>]. However, only 11,134 jury trials were completed that same year. U.S. COURTS, JUDICIAL FACTS AND FIGURES, U.S. DISTRICT COURTS—CIVIL AND CRIMINAL TRIALS COMPLETED Table 6.4 (Sept. 30, 2017), https://www.uscourts.gov/sites/default/files/data_tables/jff_6.4_0930.2017.pdf [<https://perma.cc/V92S-N6CB>].

153. See HASTIE ET AL., *supra* note 8, at 7.

154. David DeMatteo & Natalie Anumba, *The Validity of Jury Decision-Making Research*, in 1 PSYCHOLOGY IN THE COURTROOM: JURY PSYCHOLOGY, *supra* note 12, at 1, 2 (citations omitted); see also DEVINE, *supra* note 8, at 14–15.

155. DeMatteo & Anumba, *supra* note 154, at 2–3. The project's main findings were summarized in two publications: HARRY KALVEN, JR. & HANS ZEISEL, *THE AMERICAN JURY* (1966); HANS ZEISEL ET AL., *DELAY IN THE COURT* (1959).

156. DeMatteo & Anumba, *supra* note 154, at 2–3.

157. *Id.*

158. In the 1970s, psychology and law became an established independent field, witnessing the first joint-degree program in psychology and law at the University of Nebraska, the birth of the American Psychology-Law Society and the publication of *Law and Human Behavior*, the first journal dedicated to research in psychology and law and to this day one of the main journals in the field, as well as the publication of the first extensive

v. Illinois, the Court decided that excluding potential jurors during *voir dire* based on their reservations regarding the death penalty violated defendants' Sixth and Fourteenth Amendment rights.¹⁵⁹ Although the Court expressed strong reservations about the jury decision-making studies presented by the defendant—calling such studies “tentative and fragmentary,”¹⁶⁰ that decision showed the Court's willingness to consider such studies in its reasoning.¹⁶¹ In *Williams v. Florida*, the court again referred to jury decision-making studies when deciding whether the use of a six-member jury in criminal cases—instead of a traditional twelve-member jury—was constitutional.¹⁶² Citing six studies, the Court held that a jury of six was equally as effective as a jury of twelve and was thus constitutional.¹⁶³ *Williams* sparked researchers' interests, and several studies concerning jury size followed.¹⁶⁴ Many of these studies were later acknowledged in *Colgrove v. Battin* when the Court upheld and extended *Williams* to civil trials.¹⁶⁵ When the Court revisited the issue of jury size in *Ballew v. Georgia*—deciding

literature reviews of empirical research on juries. See James H. Davis, Robert M. Bray & Robert W. Holt, *The Empirical Study of Decision Processes in Juries: A Critical Review, in LAW, JUSTICE, AND THE INDIVIDUAL IN SOCIETY: PSYCHOLOGICAL AND LEGAL ISSUES* 326, 346 (June Louin Tapp & Felice J. Levine eds., 1977); DeMatteo & Anumba, *supra* note 154, at 4–5; Kathleen Carrese Gerbasi et al., *Justice Needs a New Blindfold: A Review of Mock Jury Research*, 84 *PSYCHOL. BULL.* 323, 340 (1977). Dennis Devine estimated that over 1,500 jury studies were published by the end of 2011. DEVINE, *supra* note 8, at 8. This number, Devine notes, is most likely an underestimation, “with perhaps as many as half of [the number of total studies conducted] never published.” *Id.*

159. 391 U.S. 510, 522–23, 528–29 (1968).

160. *Id.* at 517.

161. See Shari Seidman Diamond, *Illuminations and Shadows from Jury Simulations*, 21 *LAW & HUM. BEHAV.* 561, 568 (1997). The use of social science studies by the U.S. Supreme Court goes back at least as early as *Brown v. Board of Education*. See 347 U.S. 483, 494, 494 n.11 (1954). However, the use of studies specific to jury decision-making is more recent.

162. 399 U.S. 78, 135 (1970).

163. *Id.* at 86, 101–03; see also Robert J. MacCoun, *Experimental Research on Jury Decision-Making*, 30 *JURIMETRICS J.* 223, 228 (1989); David L. Suggs, *The Use of Psychological Research by the Judiciary: Do the Courts Adequately Assess the Validity of the Research?*, 3 *LAW & HUM. BEHAV.* 135, 144–45 (1979).

164. See DEVINE, *supra* note 8, at 17. Other Supreme Court decisions at around the same time also sparked interest from psychologists. One such example includes the acceptability of non-unanimous jury verdicts in restricted cases. See, e.g., *Apodaca v. Oregon*, 406 U.S. 404, 413–14 (1972); *Johnson v. Louisiana*, 406 U.S. 356, 380 (1972).

165. 413 U.S. 149, 158–60 (1973); see also Dennis J. Devine et al., *Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups*, 7 *PSYCH., PUB. POL'Y & L.* 622, 668 (2001). After *Colgrove*, the scientific community heavily criticized four of the studies cited by the Supreme Court. See Suggs, *supra* note 163, at 145.

juries smaller than six people violate defendants' Sixth and Fourteenth Amendments rights—it cited social science research indicating that deliberation in groups smaller than six people increases the chance of biases.¹⁶⁶

Many other topics besides jury selection and jury size became the subject of empirical research about jury decision-making. For instance, following the Supreme Court's upholding of capital punishment in *Gregg v. Georgia*,¹⁶⁷ many researchers turned their attention to death penalty cases, particularly focusing on the possibilities of racial biases.¹⁶⁸ Part of that research figured prominently in *McCleskey v. Kemp*, when the Court decided that the available data on existing racial biases in death penalty cases in Georgia was not enough to overturn a death sentence because the defendant had not shown that "purposeful discrimination 'had a discriminatory effect' on him."¹⁶⁹ Research on jury decision-making continued to accelerate into the first decades of the twenty-first century.¹⁷⁰

A key element of the newer research is that after federal and state governments prohibited the recording or observation of jury deliberations, studies since 1970 have focused mostly on simulated, or mock, trials.¹⁷¹

166. 435 U.S. 223, 234–38, 245 (1978); see also William C. Thompson, *Research on Jury Decision Making: The State of the Science*, in *INDIVIDUAL AND GROUP DECISION MAKING: CURRENT ISSUES* 203, 205 (N. John Castellan, Jr., ed., 1993); Suggs, *supra* note 163, at 144, 146–47.

167. 428 U.S. 153, 169 (1976).

168. See generally, e.g., David C. Baldus, et al., *Racial Discrimination and the Death Penalty in the Post-Furman Era: An Empirical and Legal Overview, with Recent Findings from Philadelphia*, 83 *CORNELL L. REV.* 1638 (1998); David C. Baldus, et al., *Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience*, 74 *J. CRIM. L. & CRIMINOLOGY* 661 (1983).

169. 481 U.S. 279, 292–93, 313 (1987) (quoting *Wayte v. United States*, 470 U.S. 598, 608 (1985)).

170. That period saw the publication of influential works. See generally, e.g., JEFFREY ABRAMSON, *WE, THE JURY: THE JURY SYSTEM AND THE IDEAL OF DEMOCRACY* (1994); STEPHEN J. ADLER, *THE JURY: TRIAL AND ERROR IN THE AMERICAN COURTROOM* (1994); VALERIE P. HANS & NEIL VIDMAR, *JUDGING THE JURY* (1986); HASTIE ET AL., *supra* note 8; RANDOLPH N. JONAKAIT, *THE AMERICAN JURY SYSTEM* (2003); SAUL M. KASSIN & LAWRENCE S. WRIGHTSMAN, *THE AMERICAN JURY ON TRIAL: PSYCHOLOGICAL PERSPECTIVES* (Barbara A. Bodling ed., 1988); NEIL VIDMAR & VALERIE P. HANS, *AMERICAN JURIES: THE VERDICT* (2007). For a good summary in the law review literature, see generally Lillquist, *supra* note 52.

171. See Devine et al., *supra* note 165, at 625–28, 670, 677, 684. The expression simulated trials actually hide different dimensions of research, such as sample employed—undergraduates or former jurors; sample size; trial presentation medium—written description of facts, written summaries, transcripts, audio or video-taped testimonies, live re-enactments; trial settings—classroom, actual courtrooms, mock courtroom, laboratories; trials elements—*voir dire*, opening and closing statements, witnesses, jury instructions, jury deliberations; and measurement of dependent variable—dichotomous, guilty or not-guilty, or continuous, estimation of probability of guilt. However, due to financial, practical, legal, and ethical constraints, most simulated trials tend to be composed of small samples of undergraduate

One benefit of this method is that researchers can systematically and predictably manipulate variables while excluding unimportant or confounding variables.¹⁷² However, whatever is gained in “internal validity”—the degree to which accurate conclusions can be drawn about what causes what or the ability of a study to rule out alternative explanations—is lost in “external validity”—the degree to which observed effects or relationships can be expected to hold in other settings or the extent to which results generalize.¹⁷³ This has led the legal profession and courts, with their high premium on external validity, to be skeptical of jury decision-making studies as well as on other types of psychological research on law-related topics.¹⁷⁴ In other words, there is a concern that studies involving undergraduate college students in simulated trials consisting almost entirely of short written summaries do not yield relevant or predictive information for real world application.

Although it is important to recognize the limitations of research, the answer to the question of external validity is by no means straightforward. In fact, supporters of jury decision-making research often argue that criticism against such research is often accompanied by a lack of understanding about methodology.¹⁷⁵ They argue that the lack of realism of simulated studies

students who are asked to read a brief written summary of facts of a case in a laboratory setting. See Joel D. Lieberman & Bruce D. Sales, *Jury Instructions: Past, Present, and Future*, 6 PSYCHOL., PUB. POL’Y & L. 587, 589 (2000); Tara L. Mitchell, et al., *Racial Bias in Mock Juror Decision-Making: A Meta-Analytic Review of Defendant Treatment*, 29 LAW & HUM. BEHAV. 621, 624 (2005). This makes findings of jury simulations less relevant to actual cases, which are notoriously lengthier and more complex.

172. See, e.g., Thompson, *supra* note 166, at 204, 207.

173. DEVINE, *supra* note 8, at 10 (emphasis omitted); Devine et al., *supra* note 165, at 698.

174. See Kerri F. Dunn, *Assessing the External Validity of Jury Simulation Research: A Meta-Analysis 2* (Aug. 2002) (unpublished Ph.D. dissertation, University of Nebraska) (on file with author). For instance, in *Lockhart v. McCree*, the Supreme Court rejected fourteen out of fifteen studies presented on the basis of supposed methodological problems. 476 U.S. 162, 169–73 (1986).

175. See, e.g., Norbert L. Kerr & Robert M. Bray, *Simulation, Realism, and the Study of the Jury*, in PSYCHOLOGY AND LAW: AN EMPIRICAL PERSPECTIVE 322, 322 (Neil Brewer & Kipling D. Williams eds., 2005) (noting it is common for critics to confuse verisimilitude for ecological validity and ecological validity for external validity). There is also research showing that there is little difference between samples constituted by students and nonstudents. See, e.g., Brian H. Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?*, 23 LAW & HUM. BEHAV. 75, 75, 78–81 (1999) (citations omitted); Joel D. Lieberman & Bruce D. Sales, *What Social Science Teaches Us About the Jury Instruction Process*, 3 PSYCHOL., PUB. POL’Y & L. 589, 617, 623, 633 (1997) (citations omitted).

does not necessarily translate into a lack of external validity, nor does it indicate that the studies' findings are any less useful.¹⁷⁶ Despite the legal profession's skepticism, research about jury decision-making might contribute to a better understanding of how jurors make decisions.¹⁷⁷ In fact, the fields of psychology and law have already made substantial progress in many interesting and relevant topics for the legal profession. Examples of topics include eyewitnesses,¹⁷⁸ pretrial publicity,¹⁷⁹ jury selection,¹⁸⁰ false convictions,¹⁸¹ and jury instructions.¹⁸²

Here, I want to highlight a few key findings from jury decision-making research that suggests that when deciding cases, jurors are not indifferent to error distribution across different situations. A group of studies focuses on the "severity-leniency hypothesis."¹⁸³ This hypothesis "asserts that jurors . . . are less willing to risk convicting an innocent person as the negative consequences"—the costs—associated with a guilty verdict increase.¹⁸⁴ In other words, there is a tradeoff between the perceived severity of the charge, or of the punishment, and the likelihood of conviction: the more severe the crime, the lower the probability of a guilty verdict. Although not explicitly about variations in the standards of proof, these studies suggest that when the charges or prescribed penalties are more severe, jurors reduce the likelihood of mistakenly punishing an innocent man by requiring more

176. DeMatteo & Anumba, *supra* note 154, at 16; *see also* Kerr & Bray, *supra* note 175, at 322–23, 358; Thompson, *supra* note 166, at 204–05, 215.

177. *See, e.g.*, HASTIE ET AL., *supra* note 8, at 238; Thompson, *supra* note 166, at 204, 207 (finding that use of simulated settings does not affect juror decision-making and that the manner in which researchers present the trial to mock jurors rarely influences the outcome); Bornstein, *supra* note 175, at 75, 78–81 (citations omitted).

178. *See generally, e.g.*, Howard E. Egeth, *What Do We Not Know About Eyewitness Identification?*, 48 AM. PSYCHOLOGIST 577 (1993).

179. *See generally, e.g.*, Christina A. Studebaker et al., *Studying Pretrial Publicity Effects: New Methods for Improving Ecological Validity and Testing External Validity*, 26 LAW & HUM. BEHAV. 19 (2002).

180. *See generally, e.g.*, JOEL D. LIEBERMAN & BRUCE D. SALES, SCIENTIFIC JURY SELECTION (2007).

181. *See generally, e.g.*, Saul M. Kassin & Gisli H. Gudjonsson, *The Psychology of Confessions: A Review of the Literature and Issues*, 5 PSYCHOL. SCI. PUB. INT. 33 (2004).

182. *See generally, e.g.*, Lieberman, *supra* note 12.

183. DEVINE, *supra* note 8, at 79–80 (emphasis omitted). Norbert Kerr makes a distinction between a hypothesis, according to which the more severe the penalty prescribed for an offense, the more evidence of guilt necessary for conviction—which he calls the "severity-criterion hypothesis"—and the more severe the penalty prescribed for an offense, the lower the likelihood of conviction—which he calls the "severity-leniency hypothesis." Norbert L. Kerr, *Severity of Prescribed Penalty and Mock Jurors' Verdicts*, 36 J. PERSONALITY & SOC. PSYCHOL. 1431, 1431–32 (1978). In this Article, I will use the expression *severity-leniency hypothesis* to refer to both hypotheses interchangeably.

184. DEVINE, *supra* note 8, at 79–80; Kerr, *supra* note 183.

conclusive evidence of guilt.¹⁸⁵ But this is just another way of saying that when the charge or prescribed penalty is more severe, jurors are more willing to raise the de facto standard of proof.

Many in the legal profession widely accept the validity of the severity-leniency hypothesis, even if only anecdotally.¹⁸⁶ In their seminal 1966 work on juries, Kalven and Zeisel asked a large group of judges to explain why they believed jurors in their courts had reached certain verdicts.¹⁸⁷ Often, judges speculated that a jury acquittal was the consequence of the jurors' feeling that the possible punishment was simply too severe.¹⁸⁸

Empirical studies do not univocally confirm or rebut the severity-leniency hypothesis,¹⁸⁹ but available data suggests that the seriousness of the charge the defendant faces affects verdicts. One study presented 227 mock jurors with a homicide case and had them choose a verdict from different combinations of the following charges: first-degree murder, second-degree murder, and manslaughter.¹⁹⁰ Some jurors had to choose between first-degree and second-degree murder, others between first-degree murder and manslaughter, others between second-degree murder and manslaughter, and so on for all seven possible combinations of the three charges plus a control condition where

185. It is important to note, however, that this strategy for reducing the risk of a false conviction would simultaneously increase the risk of a false acquittal. This means that if jurors perceive the costs of a false acquittal as sufficiently large, they might well not use this strategy. For example, if the defendant was charged as a psychopathic killer, the high cost of setting a guilty person free might restrain one from raising one's decision-making threshold. See KALVEN & ZEISEL, *supra* note 155, at 306–07.

186. See, e.g., Johannes Andenaes, *The General Preventive Effects of Punishment*, 114 U. PA. L. REV. 949, 970 (1966); John F. Galliher et al., *Nebraska's Marijuana Law: A Case of Unexpected Legislative Innovation*, 8 LAW & SOC. REV. 441, 446–47, 453 (1974).

187. KALVEN & ZEISEL, *supra* note 155, at 287–88.

188. *Id.* at 306–07.

189. See, e.g., DEVINE, *supra* note 8, at 82 (citing Jonathan L. Freedman et al., *Severity of Penalty, Seriousness of the Charge, and Mock Jurors' Verdicts*, 18 LAW & HUM. BEHAV. 189, 202 (1994)).

190. Neil Vidmar, *Effect of Decision Alternatives on the Verdicts and Social Perceptions of Simulated Jurors*, 22 J. PERSONALITY & SOC. PSYCHOL. 211, 215–17 (1972) (finding a greater percentage ($p < .001$) of innocent verdicts when mock jurors were presented with the alternatives “not guilty” or guilty of “first-degree murder” (54%) than when allowed to consider the less severe charges of “manslaughter” or “second-degree murder” (8%)). However, the author himself interpreted his results with caution because the trials confounded the severity of charges with evidentiary requirements for conviction. See *id.* at 216. For instance, charges of first-degree murder required the jury to find intention to kill and premeditation, something not necessary for a manslaughter charge. *Id.*

no verdict alternative was given.¹⁹¹ Jurors were less likely to convict when the least-serious charge—manslaughter—was not listed among the possibilities¹⁹²—a result consistent with the severity-leniency hypothesis. More specifically, “under conditions of restricted decision alternatives, the more severe the degree of guilt associated with the least severe guilt alternative, the greater were the chances of obtaining a not guilty verdict.”¹⁹³ Subsequent studies involving mock trials similarly found an inverse trend between jury verdicts and sanction severity.¹⁹⁴

191. *Id.*

192. *Id.*

193. *Id.* at 215.

194. See generally, e.g., Phoebe C. Ellsworth & Lee Ross, *Public Opinion and Judicial Decision Making: An Example from Research on Capital Punishment*, in CAPITAL PUNISHMENT IN THE UNITED STATES 152 (Hugo Adam Bedau & Chester M. Pierce eds., 1976) (finding that subjects answered positively to the question of whether they would need more evidence to vote guilty if the penalty was death instead of life imprisonment); V. Lee Hamilton, *Obedience and Responsibility: A Jury Simulation*, 36 J. PERSONALITY & SOC. PSYCHOL. 126 (1978) (finding that for a subset of cases considered, restricting the available verdict options lowered convictions); Martin F. Kaplan & Sharon Krupa, *Severe Penalties Under the Control of Others Can Reduce Guilt Verdicts*, 10 LAW & PSYCHOL. REV. 1 (1986) (finding penalty severity interacted with several other variables in a highly complex fashion); Kalman J. Kaplan & Roger I. Simon, *Latitude and Severity of Sentencing Options, Race of the Victim and Decisions of Simulated Jurors: Some Issues Arising from the “Algiers Motel” Trial*, 7 LAW & SOC. REV. 87 (1972) (finding that the severity of punishment associated with the guilty verdict is inversely related to the percentage of guilty decisions); Kerr, *supra* note 183 (“Increasing the severity of the prescribed penalty for an offence resulted in an adjustment of subjects’ conviction criteria such that more proof of guilt was required for conviction and this resulted in a reduced probability of conviction.”); Chantal Mees Koch & Dennis J. Devine, *Effects of Reasonable Doubt Definition and Inclusion of a Lesser Charge on Jury Verdicts*, 23 LAW & HUM. BEHAV. 653 (1999) (“Juries with the option to convict on a lesser charge produced more overall convictions than juries receiving only the primary charge, but only when ‘reasonable doubt’ was undefined.”); William C. McComas & Mark E. Noll, *Effects of Seriousness of Charge and Punishment Severity on the Judgments of Simulated Jurors*, 24 PSYCHOL. REC. 545 (1974) (manipulating the seriousness of the charge and the severity of the penalty and finding that only the former, not the later, produced a significant effect on jury verdicts); Rita James Simon & Linda Mahan, *Quantifying Burdens of Proof: A View from the Bench, the Jury, and the Classroom*, 5 LAW & SOC. REV. 319 (1971) (finding that the lower the penalty for an offense, the lower the probability that subjects thought defendant had committed the crime he was being charged with). But see James H. Davis et al., *Victim Consequences, Sentence Severity, and Decision Processes in Mock Juries*, 18 ORGANIZATIONAL BEHAV. & HUM. PERFORMANCE 346 (1977) (finding that manipulating the severity of the penalty associated with a rape charge using six-person mock juries has no effect on individual judgments of guilt or on jury verdicts); Freedman et al., *supra* note 189 (finding an effect for charge seriousness and no effect for penalty severity in an initial study, but the effect of charge seriousness disappeared when the strength of evidence was held constant, but still arguing that there are reasons to believe that severity of penalty might affect real—as opposed to simulated—juries).

Field studies also offer some support for the severity-leniency hypothesis. One study analyzed archival data from 201 jury trials in Indianapolis and found a negative relationship between crime severity and the likelihood of conviction when enough variables were taken into account.¹⁹⁵ Another study looked into 206 criminal trials in Utah and found a correlation—albeit weak—between the seriousness of the charge and the jury’s verdict.¹⁹⁶ These results suggest broader implications than the effects of severity of charges or penalty. Any factor that can alter a juror’s perception of costs associated with possible errors—for example, conditions of imprisonment, attitudes towards the defendant, probability of recurrence—may influence the juror’s decision-making threshold for conviction, that is the juror’s de facto standard of proof.

The literature about juries in civil cases is less developed than its criminal counterpart. Still, existing research has pointed to key variables that exert verifiable influence over the frequency and magnitude of damage awards.¹⁹⁷ One such variable is the reprehensibility of the parties’ conduct, particularly the defendant. This tends to affect the amount of damages juries award. Studies also show a bias against deep pocket defendants. In a study analyzing over twenty years of verdicts in Cook County, Illinois, researchers found juries were more likely to award more money to plaintiffs when corporations appeared as defendants.¹⁹⁸ Similar results were found in mock trial studies.¹⁹⁹

195. Martha A. Myers, *Rule Departures and Making Law: Juries and Their Verdicts*, 13 LAW & SOC. REV. 781, 785, 795 (1979).

196. See Carol Werner et al., *The Impact of Case Characteristics and Prior Jury Experience on Jury Verdicts*, 15 J. APPLIED SOC. PSYCHOL. 409, 409–10 (1985). But see Dennis J. Devine et al., *Strength of Evidence, Extraevidentiary Influence, and the Liberation Hypothesis: Data from the Field*, 33 LAW & HUM. BEHAV. 136, 145–46 (2009) (finding real jurors were actually more willing to convict under lower levels of certainty when faced with the prospect of errantly returning a violent perpetrator to their community).

197. See EDIE GREENE & BRIAN H. BORNSTEIN, DETERMINING DAMAGES: THE PSYCHOLOGY OF JURY AWARDS 25 (2003). For a skeptical view of civil jury verdicts, see CASS R. SUNSTEIN ET AL., PUNITIVE DAMAGES: HOW JURIES DECIDE 248–49 (2002) (describing civil jury verdicts as essentially groundless).

198. AUDREY CHIN & MARK A. PETERSON, DEEP POCKETS, EMPTY POCKETS: WHO WINS IN COOK COUNTY JURY TRIALS 43 (1985).

199. See, e.g., Valerie P. Hans & M. David Ermann, *Responses to Corporate Versus Individual Wrongdoing*, 13 LAW & HUM. BEHAV. 151, 157, 163 (1989) (finding juries were more likely to award plaintiffs higher damages when the defendant was referred to as “Jones Corporation” than as “Mr. Jones”); Jennifer K. Robbennolt, *Punitive Damage Decision Making: The Decisions of Citizens and Trial Court Judges*, 26 LAW & HUM. BEHAV. 315, 315 (2002) (finding judge and jury punitive damage awards “were influenced by wealth of the defendant”). But see VALERIE P. HANS, BUSINESS ON TRIAL: THE CIVIL JURY

Although not explicitly related to variations in the standard of proof, these studies suggest jurors are not indifferent between errors in favor of plaintiffs or defendants depending on the conduct involved and who is being sued.

Reactions to results like these have been mixed. At times, they take the form of criticizing the findings or methodologies of the studies.²⁰⁰ Common critiques include the fact that most subjects are students²⁰¹ and that researchers provide a short supply of information for subjects to make estimates, presumably less than real trials.²⁰² Other reactions to research on jury decision-making take the form of proposals to reform the legal system. Some have suggested instruments to increase the precision and rigor of standards such as beyond reasonable doubt.²⁰³ Others have advocated for changes to jury instructions, ranging from more precise instruction to doing away with the court's definition of reasonable doubt altogether.²⁰⁴

Albeit different, these proposals share a common feature: they all assume something is broken and needs to be fixed. More specifically, these suggestions assume there can only be one level of the standard of proof in criminal cases

AND CORPORATE RESPONSIBILITY 216–17 (2000) (describing pro-plaintiff juries and anti-business juries as myths); Robert J. MacCoun, *Differential Treatment of Corporate Defendants by Juries: An Examination of the “Deep-Pockets” Hypothesis*, 30 LAW & SOC'Y REV. 121, 121 (1996) (“While juries do appear to treat corporations differently, the explanation may have more to do with citizens’ views about the special risks and responsibilities of commercial activity.”).

200. See, e.g., Lillquist, *supra* note 52, at 112.

201. The claim that students are representative of groups of citizens called for jury duty is controversial. Compare Jonathan D. Casper & Kennette M. Benedict, *The Influence of Outcome Information and Attitudes on Juror Decision Making in Search and Seizure Cases*, in INSIDE THE JUROR: THE PSYCHOLOGY OF JUROR DECISION MAKING 65, 78 (Reid Hastie ed., 2004) (finding “no substantial differences in the patterns of awards” between citizens called for jury duty and students), with Robert J. MacCoun & Norbert L. Kerr, *Asymmetric Influence in Mock Jury Deliberation: Jurors’ Bias for Leniency*, 54 J. PERSONALITY & SOC. PSYCHOL. 21, 21 (1988) (finding differences), and Simon & Mahan, *supra* note 194, at 322.

202. Lillquist, *supra* note 52, at 115.

203. See, e.g., Jon O. Newman, *Beyond “Reasonable Doubt,”* 68 N.Y.U. L. REV. 979, 979 (1993); Robert C. Power, *Reasonable and Other Doubts: The Problem of Jury Instructions*, 67 TENN. L. REV. 45, 122–23 (1999) (proposing simplifying and clarifying jury instructions); Lawrence M. Solan, *Refocusing the Burden of Proof in Criminal Cases: Some Doubt About Reasonable Doubt*, 78 TEX. L. REV. 105, 105, 147 (1999).

204. See AMIRAM ELWORK ET AL., MAKING JURY INSTRUCTIONS UNDERSTANDABLE 25–26 (1982) (describing several approaches to improve comprehensibility of jury instructions); Jessica N. Cohen, *The Reasonable Doubt Jury Instruction: Giving Meaning to a Critical Concept*, 22 AM. J. CRIM. L. 677, 678 (1995) (demonstrating why the reasonable doubt standard should be more thoroughly “defined for the jury”); Shelagh Kenney, *Fifth Amendment—Upholding the Constitutional Merit of Misleading Reasonable Doubt Jury Instructions*, 85 J. CRIM. L. & CRIMINOLOGY 989, 1025–27 (1995) (arguing for improvement in comprehensibility of reasonable doubt instructions). But see Note, *Reasonable Doubt: An Argument Against Definition*, 108 HARV. L. REV. 1955, 1955, 1972 (1995) (“[C]ourts should not attempt to define the . . . reasonable doubt [standard] to juries.”).

and most civil cases. Any variation is an anomaly that should be fought against. This Article is intended to combat that assumption. The perceived anomaly should instead be the norm. More importantly, data about how juries decide cases strongly suggests our reality is already one of varying standards of proof.

VI. OBJECTIONS AGAINST AND REPLIES

This Part addresses potential objections against this Article's proposal and offers initial responses. It begins by discussing the claim that a system with varying standards of proof would lead to prohibitively high administrative costs and necessarily have unpredictable dynamic effects. Next, it examines the objection that individuals' cognitive limitations make the current system with few constant standards an optimal scheme. Lastly, it explores the objection that the current proposal is moot because the legal system *already* adjusts the distribution of error across different types of cases through legal instruments other than standards of proof.

A. High Administrative Costs and Unpredictable Effects

Ronald Allen and Alex Stein raised a series of objections to Kaplow's analysis that are equally relevant to the proposal in this Article.²⁰⁵ They started by pointing out the potentially high administrative costs associated with a system that attempts to define the optimal standard of proof based on a welfare analysis.²⁰⁶ This is a significant concern about any policy proposal and ought to be taken seriously. I argue, however, that objectors tend to overestimate this point.

A related objection concerns the difficulty of predicting the dynamic effects of a system with varying standards of proof. Let us pretend we were able to meet all the required administrative costs by focusing our efforts on changing the applicable standard for one category of cases, such as unpaid overtime claims. We have also successfully listed all expected social and

205. See Allen & Stein, *supra* note 9, at 582 (“To operationalize Kaplow’s proposal requires policymakers to articulate all these categories, along with all other forms of human activity, and gather dependable information about the mix of harms and benefits associated with each category. This is outlandish.”); see also CLERMONT, *supra* note 9. But see Kaplow, *Burden of Proof*, *supra* note 4, at 772 n.59 (“[T]he information requirements for determination of the optimal evidence threshold do not differ greatly from those for the determinants for the preponderance rule or other rules based on ex post likelihoods.”).

206. See Allen & Stein, *supra* note 9, at 580–81.

private harms and benefits associated with the acts as well as additional fairness and distributive considerations, and we have settled on a level of standard of proof we expect to be socially acceptable. According to objectors, the problem is that once a standard is in place, people will modify their behavior to conform to the new standard. We should hope so, because deterrence operates through behavioral changes. However, the objection persists; it is hard to predict the new equilibrium after people have adjusted to the new standard.²⁰⁷ We would have to be able to predict avoidance and exploitation efforts and continuously adjust standards. But, again, “[t]he amount of information that these [constant] adjustments will require [would be] unrealistic.”²⁰⁸ For objectors, the twin problems of high costs and unpredictable dynamic effects will most likely offset any potential gain following from a system with varying standards of proof.

These two related objections touch on a more general point about the function of standards as mechanisms for distributing errors. According to some evidence scholars, standards simply *cannot* serve this function.²⁰⁹ This is because, to guarantee a particular error distribution occurs, we need information concerning, among other things, the distribution of truly guilty and truly innocent people who go to trial, as discussed above. For example, imagine we have fifty defendants facing criminal charges; suppose forty are innocent and ten are guilty.²¹⁰ With a standard of proof of 0.9, we can predict that the jurors will wrongly convict about 10% of the innocent, which means a total of four.²¹¹ If we design our standard to distribute errors according to the idea that we should let ten guilty people go free for every innocent we condemn, we would expect to let forty guilty people go free. However, even if we let *all* of the guilty free, we would still fall short of our targeted error ratio of false acquittals to false convictions.²¹² This suggests that just defining the standard one way or the other offers no guarantee that the desired error ratio will be verified. We need information

207. Ronald J. Allen, *Rationality and the Taming of Complexity*, 62 ALA. L. REV. 1047, 1063–66 (2011) [hereinafter Allen, *Rationality*].

208. Allen & Stein, *supra* note 9, at 583.

209. See, e.g., LAUDAN, *supra* note 11, at 73. See generally Allen, *Juridical Proof*, *supra* note 11; Allen, *Rationality*, *supra* note 207; Pardo, *supra* note 11; Zoë Johnson King, *The Trouble with Standards of Proof* (Apr. 15, 2016) (unpublished manuscript) (on file with author).

210. See LAUDAN, *supra* note 11, at 73–74.

211. The standard of proof in this example is represented probabilistically for heuristic purposes only. For a critique of that approach to standards, see *supra* Part II.

212. See LAUDAN, *supra* note 11, at 74 (arguing in favor of a ratio of true acquittals to false convictions instead of false acquittals to false convictions).

about the world. The problem is that this information is difficult, if not impossible, to obtain.²¹³

There are at least two possible reactions to this difficulty. We can surrender the idea that standards can serve as mechanisms to distribute errors. Alternatively, we can change our understanding of how standards fulfill this function. Here, I explore this second option. Only after exhausting our efforts can we concede this very plausible and important role we assign to standards.

One way to change our understanding of how standards of proof can fulfill their function of distributing errors is by lowering our expectations about our capacity to verify whether we have reached an exact error-distribution. An analogy might be helpful here. When discussing legal rules, Frederick Schauer treats them as “prescriptive generalizations.”²¹⁴ The desire to promote safe driving and the idea that most people when driving over fifty-five miles per hour on a street are prone to unsafe driving, allows us to generalize that *all* people driving over fifty-five miles per hour on that street should pay a fine. Because this generalization is necessarily over- and under-inclusive—there might be people who drive safely over fifty-five—the results indicated by applying a rule will, in some cases, be inferior to the results indicated by directly applying the rule’s background justifications. This suggests that rule-based decision-making will be suboptimal, failing to achieve the best result on every occasion—the result in which the justification underlying each generalization is correctly applied to each case.²¹⁵ The fact that rules are suboptimal does not mean that the procedure of deciding according to rules is suboptimal. In fact, the optimal decision procedure may *not* be the one aimed at producing the optimal result in *every* case. A procedure aimed at optimality in every individual case may produce worse results in the aggregate. In this sense, rules are second-best solutions.

213. This challenge increases once we try to include in the analysis the effects of other legal institutions on the error-distribution function of standards. Given their ubiquity in contemporary adjudicatory systems, plea bargains and settlements stand out as important examples to be analyzed. One way in which pleas may impact error-distributions is by changing the distribution of truly guilty and truly innocent who go to trial. But how can we measure those effects? Are they the same for all types of cases? Or should we expect the effect of pleas for the distribution of truly guilty and truly innocent that go to trial to differ depending on the type of crime and type of defendant involved?

214. FREDERICK SCHAUER, *PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING IN LAW AND IN LIFE* 25–27 (1991).

215. *Id.* at 100–02. See generally Frederick Schauer, *In Defense of Rule-Based Evidence Law—and Epistemology Too*, 5 *EPISTEME* 295 (2008).

I suggest that something similar is true of standards of proof. We can see them as generalizations with a proper underlying justification. In the case of standards, this justification is the goal of achieving a given error distribution. Similar to legal rules, standards can be seen as *suboptimal* generalizations. But their suboptimality does not come from over- and under-inclusiveness. Rather, it comes from constraints on whether we can verify that the targeted error distribution has been achieved. These constraints are both epistemic—we have imperfect knowledge about the distribution of truly guilty and truly innocent that go to trial—and formal—we cannot wait for relevant data to decide cases, nor can we revisit them if future data suggests something different from what was thought when the decision was made.

It is important to note that the generalizations that standards represent are not arbitrary; they are not spurious generalizations. First, we have plausible intuitions about the expected distribution of truly guilty and truly innocent people as well as the potential effects that establishing the standards at different levels can cause. To the extent that these intuitions are plausible and coherent with our background beliefs, they are *prima facie* epistemically justified.²¹⁶ Second, we can periodically stop to re-evaluate the standards. If we notice any great deviation between our reality and our ambitions, we can make the appropriate adjustments.²¹⁷

More generally, this type of academic debunking can be aimed at any policy analysis. The problem is that discarding a policy proposal as mere speculation until further empirical information becomes available imposes unreasonable constraints on policy-oriented scholarship. It might be premature to dismiss recommendations simply for lack of *sufficient data*. Surely, any model becomes increasingly dangerous as it ignores existing data.

216. This is based on the idea that the fact that a person has an intuition provides that person with *prima facie* epistemic justification in believing the content of the intuition. In other words, the fact that a particular proposition is *intuitive* provides us with a *prima facie* justification for believing in it. For a more formal and extensive discussion about the role of rational intuition in the epistemic justification for some of our belief about the empirical world see, for example, LAURENCE BONJOUR, IN DEFENSE OF PURE REASON: A RATIONALIST ACCOUNT OF A PRIORI JUSTIFICATION 100–06 (1998). *But see generally* R. M. Hare, *Critical Study: Rawls' Theory of Justice—II*, 23 PHIL. Q. 241 (1973) (arguing that a set of coherent beliefs that have no independent initial credibility cannot produce epistemic or moral justification because coherent fictions are still only fictions).

217. One might hold onto the seemingly plausible assumption that the justice system as a whole works to create more evidence against truly guilty defendants than against truly innocent defendants. *See, e.g.*, Bell, *supra* note 19, at 560–62. The hope is that we would then be able to estimate the prior allocation of truly guilty and truly innocent defendants at trial and, consequentially, calibrate the error-distribution based on the level of the standard of proof. This strategy is doomed to fail, however. It is not enough for a complete analysis of error distribution that the justice system as a whole works to create more evidence against guilty defendants than against innocent defendants. It must do so in a consistent and predictable manner. Otherwise, no theoretical analysis can be successful.

However, absent good substitutes, models that include currently unverified, but verifiable, assumptions about our empirical world can provide helpful starting points. Most importantly, these starting points need not be spurious assumptions. We can build our starting points on top of plausible intuitions about a range of empirical phenomena, including the expected distribution of truly guilty and truly innocent people and the number of individuals from each group that goes to trial with a given set of facts. These intuitions can give us *prima facie* epistemic justification for believing their content. We should not remain uncritical, however. We can—and should—pause from time to time to re-evaluate our beliefs and practices. If we note any great deviation between reality and our ambitions, we should make the appropriate adjustments and move forward. It makes little sense to substitute some other model because it was easier to employ when that substitute essentially has no relationship to our objectives.²¹⁸ In this sense, we are like Neurath's sailors, who have to rebuild their ship on the open sea.²¹⁹

One last point about these objections concerns the underlying institutional details of this Article's proposal. One configuration that might exacerbate these difficulties is gathering all that information to design the optimal standard and making relevant moral and political decisions on a *case-by-case* basis.²²⁰ Although the optimal level of generality of standards of proof is beyond the scope of this Article, it is reasonable to suppose that—given the amount of information needed to perform this analysis combined with the fact that this information is likely to vary between different cases—requiring courts to define the optimal standard based on a mix of welfare, fairness, and distributive considerations for every case would be prohibitively costly. This would be asking courts to perform tasks they might not be accustomed or equipped to perform properly.²²¹

218. The alternative with the lowest informational requirements would be to simply toss a coin, but the ease of this method gives little reason in its favor.

219. See OTTO NEURATH, *PHILOSOPHICAL PAPERS 1913–1946*, at 92 (Robert S. Cohen & Marie Neurath eds. & trans., 1983).

220. For arguments in defense of a case-by-case determination of standards of proof, see, for example, Lillquist, *supra* note 52, at 147–62; Dale A. Nance, *Evidential Completeness and the Burden of Proof*, 49 *HASTINGS L.J.* 621, 626–32 (1998). A case-by-case proposal also exacerbates the difficulty behind individuating and counting standards. Thank you to William Twining for elucidation on this point.

221. See Allen & Stein, *supra* note 9, at 601; see also Tribe, *supra* note 30, at 1384–85 (“[O]ne will expect the lawmaker rather than the factfinder to use a model such as the one Kaplan and Cullison propose, and one will define the decision problem to be solved not as the one-shot problem of fixing a standard of proof for a particular trial with four

Moreover, requiring judges and jurors to define the standard based on welfare, fairness, or distributive considerations might increase the possibility of similarly situated individuals being subjected to different standards of proof. This, in turn, might raise equal protection obstacles. The recommendation that we view standards as prescriptive generalizations suggests a system not very much unlike ours, which assigns standards *ex ante* by elected representatives, instead of on a case-by-case basis. The main difference is that we could presumably track societal choices about preferable error-distribution in different circumstances more closely. This alternative would impose lower informational requirements on decision-makers in comparison to a case-by-case alternative because decisions are made in the aggregate by decision-makers institutionally equipped with the necessary tools to perform those tasks. Most importantly, such a system would be preferable, not because these institutions are more likely to get decisions right but because this is a democratically fair way to decide politically and morally sensitive issues that reasonable people disagree about.²²²

B. Cognitive Limitations

Kevin Clermont has argued that findings from cognitive psychology literature give us good reasons to limit the number of standards of proof into our well-known trinity.²²³ Clermont recognizes that a system of varying standards is conceivable, and perhaps even theoretically, desirable.²²⁴ But he quickly warns about practical dangers. Specifically, he refers to research about individuals' cognitive limitations to judge probabilities.²²⁵ Studies suggest that judgments about probabilities are carried out in terms of a very limited set of broad and fuzzy categories such as more likely than

possible outcomes, but as the much larger problem of establishing such standards for the trial system as a whole.”). The Supreme Court argued similarly:

[T]his court never has approved case-by-case determination of the proper standard of proof for a given proceeding. Standards of proof, like other “procedural due process rules, are shaped by the risk of error inherent in the truth-finding process as applied to the generality of cases, not the rare exceptions.” [Because] the litigants and the factfinder must know at the outset of a given proceeding how the risk of error will be allocated, the standard of proof necessarily must be calibrated in advance.

Santosky v. Kramer, 455 U.S. 745, 757 (1982) (emphasis omitted) (quoting Mathews v. Eldridge, 424 U.S. 319, 344 (1976)).

222. See DWORKIN, *supra* note 5, at 24, 27.

223. See CLERMONT, *supra* note 9, at 113; see also Kevin M. Clermont, *Procedure’s Magical Number Three: Psychological Bases for Standards of Decision*, 72 CORNELL L. REV. 1115, 1146 (1987).

224. See CLERMONT, *supra* note 9, at 112.

225. See *id.* at 107.

not, high probability, and almost certainty.²²⁶ For Clermont, the existing coarsely gradated standards of proof in law track these fuzzy categories.²²⁷ This allows legal decision-makers to use their built-in cognitive apparatus while judging factual propositions in legal cases according to the three familiar standards of proof. Any deviation from those hard-wired categories would impose significant cognitive costs on individuals and increase the chance of errors. This means that any proposal attempting to change standards would have to compensate for those additional costs. Consider Clermont's words:

[L]awmakers can adequately serve the often imprecise policies underlying the standard of decision by choosing from the set of seven categories, thus moving up or down by fairly small quantum leaps rather than by unrealistically finer degrees.

....

[Any novel standard of proof] would be difficult to articulate and communicate and comprehend and apply, would risk confusion and possibly abuse by the decisionmaker, and would inevitably drift toward one of the customary standards—and that these costs more than offset any benefit of utilizing an unusual standard [of proof].²²⁸

It should not be surprising that the laypeople are not particularly good at reasoning with probabilities. Decades of psychological studies have helped to formulate, explore, and catalog the distinctive, persistent, and predictable errors in people's probabilistic reasoning, such as base-rate fallacy,²²⁹ gambler's fallacy,²³⁰ and representativeness heuristic.²³¹

This argument, however, has a special grip against probabilistic approaches to standards of proof. According to these approaches, standards of proof should be understood as probability thresholds ranging from zero to one,

226. *Id.* at 115.

227. *See generally, e.g.,* David L. Schwartz & Christopher B. Seaman, *Standards of Proof in Civil Litigation: An Experiment from Patent Law*, 26 HARV. J.L. & TECH. 429 (2013) (conducting an experiment which showed that jury instructions tended to move away from clear and convincing towards preponderance of the evidence).

228. CLERMONT, *supra* note 9, at 79–81.

229. *See, e.g.,* Amos Tversky & Daniel Kahneman, *Evidential Impact of Base Rates, in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES* 153, 153–60 (Daniel Kahneman et al. eds., 1982); Maya Bar-Hillel, *The Base-Rate Fallacy in Probability Judgments*, 44 ACTA PSYCHOLOGICA 21, 213 (1980).

230. *See, e.g.,* Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1125–30 (1974).

231. *See, e.g.,* Daniel Kahneman & Amos Tversky, *Subjective Probability: A Judgment of Representativeness*, 3 COGNITIVE PSYCHOL. 430, 430–31 (1972).

involving either objective or subjective probabilities.²³² People who subscribe to this interpretation commonly assume that preponderance of the evidence means that it is proven by a probability higher than 0.5, while beyond a reasonable doubt would require a much higher probability of conviction; in fact, it is not uncommon to hear unofficial estimates of 0.9, 0.95, or 0.99.²³³

Some evidence law scholars push for a different approach to standards, away from traditional probabilistic approaches. Under an explanation-based approach, legal fact-finding involves a determination of the best explanation of the admitted evidence rather than a determination of whether each element is proven to a specific probability.²³⁴ In a nutshell, from the fact that a given hypothesis—out of the set of hypotheses offered by the parties or constructed by the fact-finders—best explains the admitted evidence, fact-finders should infer the truth of that hypothesis and then decide the case based on that inference. Because this approach removes probabilistic reasoning as the only important consideration when dealing with standards of proof, the fact that most people are bad at reasoning with probabilities loses much of its argumentative strength.

Proponents of the explanation-based approach highlight how the approach provides a *better* description of jurors' reasoning.²³⁵ To support this, they offer evidence that what the approach asks of jurors is compatible with research on jury decision-making.²³⁶ The main finding in the studies is that, contrary to what the probabilistic approach to standards says of jurors, individuals do not usually reason with isolated pieces of evidence.²³⁷ That is, we usually do not update our beliefs based on each piece of new information

232. See, e.g., Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489, 504 (1970); David Hamer, *Probabilistic Standards of Proof, Their Complements and the Errors that Are Expected to Flow from Them*, 1 U. NEW ENG. L.J. 71, 73–74 (2004); Kaplan, *supra* note 81, at 1066; D.H. Kaye, *Clarifying the Burden of Persuasion: What Bayesian Decision Rules Do and Do Not Do*, 3 INT'L J. EVIDENCE & PROOF 1, 4–5 (1999); Jonathan J. Koehler & Daniel N. Shaviro, *Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247, 253 (1990); Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1124–25 (1977); Mike Redmayne, *Exploring the Proof Paradoxes*, 14 LEGAL THEORY 281, 281, 285 (2008).

233. See, e.g., Ronald J. Allen, *Explanationism All the Way Down*, 5 EPISTEME 320, 320–21 (2008).

234. See Nancy Pennington & Reid Hastie, *A Cognitive Theory of Juror Decision Making: The Story Model*, 13 CARDOZO L. REV. 519, 523–25 (1991).

235. See Allen, *supra* note 22, at 216.

236. See *id.*

237. See generally Nancy Pennington & Reid Hastie, *The Story Model for Juror Decision Making*, in INSIDE THE JUROR: THE PSYCHOLOGY OF JUROR DECISION MAKING, *supra* note 201, at 192; Pennington & Hastie, *supra* note 234.

we receive. Instead, we tend to construct narratives that will fit the evidence and make decisions based on the explanatory virtues of those stories.

In a recent piece, Clermont addressed the explanation-based approach to standards.²³⁸ He notes that a difficulty with this approach is that “it does not track well what the law tells its factfinders about how to proceed, and it diverges from the law by compelling the nonburdened party to choose and formulate a competing version of the truth.”²³⁹ This is not entirely accurate. The explanation approach is partly revisionary in that it pushes for an interpretation of standards that is distinct from the way many people think about those mechanisms. But, the approach is not *completely* revisionary. First, many judicial opinions explicitly reference the importance of explanatory value in legal reasoning.²⁴⁰ Second, many evidentiary mechanisms are better understood and explained if we assume they are also meant to promote explanatory value.²⁴¹ Moreover, nothing in the explanation approach requires the non-burdened party to present an alternative explanation to that presented by the party bearing the burden of persuasion. In a typical civil case, the defendant can remain silent while the plaintiff tries to convince triers that he has met his burden—even though this is almost never the best defense strategy.²⁴²

238. See generally CLERMONT, *supra* note 9.

239. *Id.* at 140.

240. The Supreme Court employed an explanatory approach in a summary-judgment case:

Neither the Court of Appeals, nor the respondents, nor the dissent provides any reason to question the city’s theory. In particular, they do not offer a competing theory, let alone data, that explain why the elevated crime rates in neighborhoods with a concentration of adult establishments can be attributed entirely to the presence of walls between, and separate entrances to, each individual adult operation.

City of Los Angeles v. Alameda Books, Inc., 535 U.S. 425, 437 (2002).

241. See, e.g., FED. R. EVID. 106 (establishing as admissible surrounding material relevant to specific testimony); FED. R. EVID. 612 (explaining if a witness relies on a piece of writing while or before testifying, that writing is admissible regardless of other exclusionary rules).

242. Even the party that does not carry the burden of persuasion almost always presents evidence in her favor. The risks of losing and the effects of preclusion create enough incentive. The fact that the burdens of production and persuasion are on the plaintiff does not mean that the defendant should rest silent. Theoretically, he can. Strategically he never should. See *How Courts Work*, AM. BAR ASSOC., https://www.americanbar.org/groups/public_education/resources/law_related_education_network/how_courts_work/defense.html [<https://perma.cc/DKA9-F9PY>].

Most importantly for this Article, an explanation-based approach to standards can offer a plausible reply to the cognitive limitation objection. Because the approach does not reduce factual decision-making in light of standards of proof to probabilistic reasoning even though probability can still play *some* role, research showing how individuals are not particularly good at reasoning with probabilities is not an insurmountable obstacle.

According to Clermont, to the extent that we lack a good reason to do things differently from what cognitive research tells us we are used to doing, we should refrain from adopting policies that have a high risk of leading decision-makers to confusion and error.²⁴³ However, we do have good reasons. Parts IV and V contain several reasons for moving to a system of varying standards of proof. Those may not be conclusive reasons, but it should be clear that, contrary to what Clermont claims, we *do* have reasons—*good* reasons—to want to do things differently. And better, we have reasons to think that the result will not be so pessimistic as the objection from cognitive limitations suggests.

C. *Already Varying De Facto Standards of Proof*

The last objection I consider questions my initial premise that our legal system assigns the same standard of proof to drastically different cases under the justification that we already accept the same error distribution in those cases. According to this objection, our legal system *already* adjusts the distribution of error across different types of cases, only it does so through legal tools other than standards of proof. Insofar as these other instruments impact how easy or difficult it is to impose liability on potential defendants, they can cause the de facto standard of proof to vary, effectively changing the error distribution.

Albeit a possibility, the strategy to pursue changes in the de facto distribution of errors through mechanisms other than standards of proof is highly questionable. In many circumstances, there will be important reasons to prefer to adjust the error distribution through changes in the applicable standard of proof. To see why, consider the following example.²⁴⁴ Suppose that an existing legal rule prohibiting certain acts thought to be harmful causes a higher number of errors by chilling benign acts. Let us refer to that set as *set X*. How do we remedy this problem? Should we narrow the existing legal rule, perhaps by exempting from liability some readily identifiable subset of *set X* (*subset x*) and leaving only the remaining *set X* minus *subset x* (remainder subset will be *subset X \ x*) subject to liability?

243. See CLERMONT, *supra* note 9, at 280.

244. The example that follows is adapted from Louis Kaplow, *Multistage Adjudication*, 126 HARV. L. REV. 1179, 1229–32 (2013).

Or is it better to increase the formal standard for *set X*, making it less likely that those benign acts will be chilled because fewer defendants will be convicted because of the higher standard, while simultaneously lowering the standard for cases in the remainder *subset X*x?

Let us start with the first strategy. Its benefits seem to depend crucially on *subset x* having two features. First, a sizable portion of the benign acts that may be subject to liability when the law targets the entire initial *set X* must be easily mistaken for acts in *subset x*, those exempted from liability under the proposed strategy. But the benign acts in *set X* cannot be mistaken for those acts in the remaining *set X*x, including those acts still subject to liability. In other words, the benign acts in *subset x*, currently exempted from liability, are indistinct from the benign acts initially subject to liability in *set X*. Second, *subset x* must not contain a high number of harmful acts. Most of the acts in *subset x* must be truly benign. If these two conditions hold, then exempting acts in *subset x* from liability might greatly reduce chilling effects without greatly undermining deterrence. This is because most of the acts exempted from liability constitute benign acts. This result makes the first strategy seem quite appealing.

The problem is that it is highly unlikely that we will ever be able to identify readily a set of acts with these features. Most of the time, the best we might do is identify a particular target subset (*subset x'*) with a slightly lower ratio of chilling costs to deterrence benefits than the remaining subset *X*x'. That is, a *subset x'* containing a higher number of benign acts compared to harmful acts that might be a plausible candidate for exemption.

There are important reasons to believe that the second strategy, namely one that increases the applicable standard of proof for cases within *x'* while lowering the standard for cases in the remainder subset (*X*x'), will be superior. One way to see why is to note that exempting from liability acts in new *subset x'* is equivalent to setting an infinitely high standard of proof for all cases involving acts in *x'*. Remember from Section IV.A that from a welfare-based analyses, the optimal standard of proof for a given situation is a function of marginal deterrence gains and chilling costs. If we were to set standards of proof aimed toward welfare maximization, the optimal standard in cases involving acts in *subset x'*, in which there are a higher number of benign acts than harmful acts, would already tend to be higher than the optimal standards in cases involving acts in remaining *subset X*x', where there is a higher number of harmful acts compared to benign acts. If implemented optimally, this is the result that the second

strategy suggests.²⁴⁵ Note, moreover, that when x' is indistinguishable from x , this result is indistinguishable from the result following the first strategy.

Let us consider briefly a situation in which there is no easily identifiable *subset* x with a slightly lower ratio of chilling costs to deterrence benefits compared to remaining *subset* $X \setminus x$. Still, lawmakers consider exempting acts in *subset* x from liability as a solution to the perception of excessive chilling of benign acts. Here, it seems that a superior strategy would be to increase the applicable standard instead of exempting those acts from liability. Increasing the applicable standard removes those cases from the legal system with the weakest level of evidential support in favor of the party carrying the burden of proof. Instead, exempting an arbitrary *subset* x from liability would remove cases that vary greatly in terms of evidential support. We would be removing from not only weak cases, which is a good thing for maximizing accuracy, but also strong cases, which is a bad thing for accuracy. With a focus on maximizing accuracy, it seems like a better strategy not to avoid cases with random levels of evidential support, but instead to concentrate on those that have the *lowest* levels.

We end up with a scenario in which either the procedural strategy is superior or equally preferable to the substantive strategy.²⁴⁶ This gives us a good reason to prefer the procedural strategy.²⁴⁷ It will usually be best to alter the error-distribution to a socially desirable ratio through the mechanism that influences that distribution—standards of proof.

245. Remember that the subset x is one in which the following two favorable conditions hold. First, a large portion of the benign acts that may be subject to liability when the law targets the entire initial set X are readily mistaken for acts in x , but not for those acts in the remaining set $X \setminus x$. Second, x does not contain very many harmful acts.

246. This does not mean that a substantive strategy will never be preferable. Procedural rules are not always optimally created and enforced, which can require substantial informational costs, as discussed above in Section VI.A. Moreover, there may be other benefits associated with a substantive strategy, such as constraining abuses of discretion and providing clearer guidance for citizens.

247. Here, one might accuse that I have failed to acknowledge an alleged substantial cost associated with the procedural strategy. The cost is related to a purported fact that fixed formal standards would have an important legitimizing function. *See, e.g.,* Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357, 1362 (1985); Tribe, *supra* note 30, at 1393. Insofar as the procedural strategy advocates for a system with varying standards, it deprives it of its legitimizing function. Even if I concede that standards might have such legitimizing goal—a questionable proposition—it is far from clear that a system with fixed standards would have that desired effect. A system with fixed standards of proof might very well have a delegitimizing effect instead. This may be because such a system incurs costs related to high chilling or loss of deterrence—because the fixed standard is far from the optimal standard from a welfare-based analyses—or because such a system is considered unfair—because it fails to give people procedures that correlate the correct importance to the risk of moral harm they face.

VII. CONCLUSION

We often hear the suggestion that, because standards of proof affect the distribution of errors in adjudication, the choice of the applicable standard in a case should reflect an assessment of the comparative weights of the errors involved. At the same time, we tell ourselves that the same standard should govern every criminal case and, virtually, every civil case. These two ideas contradict each other. If we are serious about using the socially desirable error distribution to determine the optimal standard, then we cannot plausibly also maintain that one fixed standard should govern cases in which errors for the defendant or plaintiff carry such vastly different weights. We need to abandon one of these two initial ideas. This Article argues that the idea of fixed standards should be rejected.

This rejection might seem particularly problematic because today there appears to be only three viable choices in the range of standards. But this was not always the case. Only relatively recently did courts recognize that the legal profession had reduced all conceivable possibilities into three options.²⁴⁸ Since then, we have become complacent in thinking that our limited menu exhausts all theoretical and practical possibilities. This does not have to be the case. Instead, we can—and should—adopt a system with varying standards of proof.

248. See CLERMONT, *supra* note 9, at 278; J.P. McBaine, *Burden of Proof: Degrees of Belief*, 32 CAL. L. REV. 242, 246–47 (1944); see also *Addington v. Texas*, 441 U.S. 418, 423 (1979) (“[T]he evolution of this area of the law has produced across a continuum three standards or levels of proof for different types of cases.”).

