THE YALE LAW JOURNAL

LOUIS KAPLOW

Burden of Proof

ABSTRACT. The burden of proof is a central feature of all systems of adjudication, yet one that has been subject to little normative analysis. This Article examines how strong evidence should have to be in order to assign liability when the objective is to maximize social welfare. In basic settings, there is a tradeoff between deterrence benefits and chilling costs, and the optimal proof requirement is determined by factors that are almost entirely distinct from those underlying the preponderance of the evidence rule and other traditional standards. As a consequence, these familiar burden of proof rules have some surprising properties, as do alternative criteria that have been advanced. The Article also considers how setting the proof burden interacts with other features of legal system design: the determination of enforcement effort, the level of sanctions, and the degree of accuracy of adjudication. It compares and contrasts a variety of legal environments and methods of enforcement, explaining how the appropriate proof requirements differ qualitatively across contexts. Most of the questions raised and answers presented differ in kind—as well as in result—from those in prior literature.

AUTHOR. Finn M.W. Caspersen and Household International Professor of Law and Economics, Harvard Law School, and Research Associate, National Bureau of Economic Research. I am grateful to Jody Forchheimer, Sonia Jaffe, Daniel Klerman, Steven Shavell, Matthew Stephenson, and participants in workshops at Columbia University, Harvard University, Northwestern University, Stanford University, UCLA, the University of Michigan, the University of Southern California, the University of Texas, and the University of Virginia for helpful discussions and comments; Michael Atamas, Sarah Beckerman, Noam Lerer, Daniel O'Brien, Stephen Pezzi, and Zachary Shapiro for research assistance; and Harvard Law School's John M. Olin Center for Law, Economics, and Business for financial support.

ARTICLE CONTENTS

INTRODUCTION

Systems of adjudication base outcomes on whether the strength of available evidence satisfies a designated burden of proof. Similar modes of decisionmaking are employed in many important nonlegal settings, such as when medical treatments are selected in light of test results and other diagnostics. The stringency of the proof burden determines how error is allocated between mistakes of commission—improper assignment of liability—and mistakes of omission—improper exoneration.

This Article explores how to set the evidence threshold[1] in the manner that best advances social welfare.[2] It seeks less to displace entrenched views than to fill a vacuum, for prior scholarship devotes surprisingly little attention to the rationale for how stringent proof burdens should be. In the United States, civil litigation ordinarily is governed by a preponderance of the evidence rule, under which the plaintiff must establish that it is more likely than not that the defendant is liable;[3] in criminal trials, guilt must be established beyond a reasonable doubt.[4] Neither requirement is particularly controversial, and

1. I employ the term "evidence threshold," a concept elaborated in Subsection I.A.2, to refer to the strength of the evidence required for liability. As will be noted in this Introduction and developed in Subsection II.C.1, conventional burden of proof conceptions (regarding the persuasion burden, see note 5) have a subtle, complex, and sometimes counterintuitive relationship with the minimal strength of evidence that is required for liability. This is true even though the two notions are ordinarily (implicitly) taken to be synonymous or at least related to each other in a simple, positive manner.

2. The social welfare objective—referring to the well-being of all members of society—is articulated and defended in LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE (2002). For further discussion of its propriety in the present setting, see note 42.

3. On the probabilistic interpretation of the preponderance rule, see, for example, 2 MCCORMICK ON EVIDENCE 484 (Kenneth S. Broun ed., 6th ed. 2006) ("The most acceptable meaning to be given to the expression, proof by a preponderance, seems to be proof which leads the jury to find that the existence of the contested fact is more probable than its nonexistence." (citing MODEL CODE OF EVIDENCE R. 1(3))), and David Kaye, *Naked Statistical Evidence*, 89 YALE L.J. 601, 603 (1980) (reviewing MICHAEL O. FINKELSTEIN, QUANTITATIVE METHODS IN LAW: STUDIES IN THE APPLICATION OF MATHEMATICAL PROBABILITY AND STATISTICS TO LEGAL PROBLEMS (1978)) ("A majority of courts and almost all commentators have concluded that [the preponderance of the evidence rule] is satisfied by evidence that indicates to the trier of fact that the event that must be established is more likely to have occurred than not.").

4. Interestingly, civil law jurisdictions do not apply clearly distinct proof standards in criminal and noncriminal cases, although it is not entirely obvious how their proof standards, formulated in terms of subjective belief, compare to those in the United States. For differing views, see Kevin M. Clermont & Emily Sherwin, *A Comparative View of Standards of Proof*, 50 AM. J. COMP. L. 243, 245-51 (2002); Christoph Engel, *Preponderance of the Evidence Versus*

perhaps for this reason many treatises and texts on the law of evidence and on civil and criminal procedure provide little or no explanation for these rules. By contrast, somewhat greater attention has been devoted to the production burden[5] (how much evidence a party must present in order to avoid losing by default) and tiebreaking[6] (which party wins in a civil case when the factfinder believes there is a fifty percent chance that each side is correct)—subjects that are not the focus here.

When the question is broached directly, commonly proffered justifications are as unsatisfying as they are brief. In the civil context, the primary argument for the preponderance rule seems to be a lack of any apparent reason to do otherwise, a view often expressed by noting the absence of the special features of the criminal context.[7] Some authorities go so far as to express society's

Intime Conviction*: A Behavioral Perspective on a Conflict Between American and Continental European Law*, 33 VT. L. REV. 435, 438-42 (2009); and Michele Taruffo, *Rethinking the Standards of Proof*, 51 AM. J. COMP. L. 659, 664-69 (2003). One might have expected this difference between systems to spur substantial debate, but it has not.

5. *See, e.g.*, 2 MCCORMICK ON EVIDENCE, *supra* note 3, §§ 336-38; 9 JOHN HENRY WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW §§ 2486-93 (James H. Chadbourn rev., 1981). For the classic statement distinguishing the production burden from the persuasion burden (the stringency of the burden of proof), see James B. Thayer, *The Burden of Proof*, 4 HARV. L. REV. 45 (1890), subsequently elaborated in JAMES BRADLEY THAYER, A PRELIMINARY TREATISE ON EVIDENCE AT THE COMMON LAW 353-89 (Boston, Little, Brown & Co. 1898) [hereinafter THAYER, A PRELIMINARY TREATISE]. For discussions of terminology, see note 1; BLACK'S LAW DICTIONARY 190-91 (7th ed. 1999) (noting a modern trend to use the term "burden of proof" to refer only to the persuasion burden); THAYER, A PRELIMINARY TREATISE, *supra*, at 384-89; and 21B CHARLES ALAN WRIGHT & KENNETH W. GRAHAM, JR., FEDERAL PRACTICE & PROCEDURE: FEDERAL RULES OF EVIDENCE § 5122 (2d ed. 1987) ("McCormick's oft-quoted 'assertion that "presumption" is the slipperiest member of the family of legal terms, except for its first cousin, "burden of proof"' still pops up 50 years after it was first penned." (quoting CHARLES T. MCCORMICK, HANDBOOK OF THE LAW OF EVIDENCE 639 (1954))). Many of the standard rationalizations for allocating the production burden likewise have little content. *See, e.g.*, Bruce L. Hay, *Allocating the Burden of Proof*, 72 IND. L. J. 651, 656 (1997) (presenting criticism); Fleming James, Jr., *Burdens of Proof*, 47 VA. L. REV. 51, 58-63 (1961) (criticizing some standard arguments but offering more cogent reasons); *cf.* EDMUND MORRIS MORGAN, SOME PROBLEMS OF PROOF UNDER THE ANGLO-AMERICAN SYSTEM OF LITIGATION 75-76 (1956) (criticizing parallel rationales for allocating the burden of persuasion). Some prior law and economics work, much of which emphasizes incentives to present evidence and costs of adjudication, has examined the placement of the production burden. *See, e.g.*, Hay, *supra*; Bruce L. Hay & Kathryn E. Spier, *Burdens of Proof in Civil Litigation: An Economic Perspective*, 26 J. LEGAL STUD. 413 (1997); Joel Sobel, *Disclosure of Evidence and Resolution of Disputes: Who Should Bear the Burden of Proof?*, *in* GAME-THEORETIC MODELS OF BARGAINING 341 (Alvin E. Roth ed., 1985).

6. *See infra* note 34.

7. *See, e.g.*, *In re* Winship, 397 U.S. 358, 371 (1970) (Harlan, J., concurring) ("In a civil suit between two private parties for money damages, for example, we view it as no more serious

negligible concern as to the outcome,[8] which if pressed might lead one to wonder why civil litigation should exist at all (or why the authors bother to produce elaborate texts and treatises on how it should be conducted, while others write extensively on what substantive rules should govern). It is hard to avoid the conclusion that the strong attraction of the 50% requirement is substantially attributable to its being a powerful focal point, some of its power

in general for there to be an erroneous verdict in the defendant's favor than for there to be an erroneous verdict in the plaintiff's favor." (This statement is followed almost immediately by a contrast with the criminal setting, quoted in note 10.)); Grogan v. Garner, 498 U.S. 279, 286 (1991) ("Because the preponderance-of-the-evidence standard results in a roughly equal allocation of the risk of error between litigants, we presume that this standard is applicable in civil actions between private litigants unless 'particularly important individual interests or rights are at stake.'" (quoting Herman & MacLean v. Huddleston, 459 U.S. 375, 389-90 (1983))); 2 MCCORMICK ON EVIDENCE, *supra* note 3, at 491 (stating, without further elaboration, "[m]istakes will be made and in a civil case a mistaken judgment for the plaintiff is no worse than a mistaken judgment for the defendant" (a claim that is promptly contrasted with the criminal setting, as quoted in note 10)); RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 827 (8th ed. 2011) ("In the typical civil trial, there is no basis for supposing that Type I errors (false positives, such as convicting an innocent person, or, in the civil context, erroneously finding the defendant liable) on average impose higher costs than Type II errors (false negatives, such as an erroneous acquittal or the denial of a meritorious claim). So it is enough in the usual civil case to justify a verdict for the plaintiff that the probability that his claim is meritorious exceeds, however slightly, the probability that it is not." (footnote omitted) (This statement is followed shortly by a contrast to the criminal setting, in which "Type I errors are more serious than Type II errors.")); V.C. Ball, *The Moment of Truth: Probability Theory and Standards of Proof*, 14 VAND. L. REV. 807, 816-17 (1961) ("When we come to the requirement of proof in civil actions, the case stands different. The problem then is whether there is a basis for giving a higher value to one of the two kinds of mistake, and applying a requirement which is intended to (1) increase the total number of mistakes, and (2) change the proportion and number of one kind of mistake. The majority of courts seem to have said there is not."); Mike Redmayne, *Standards of Proof in Civil Litigation*, 62 MOD. L. REV. 167, 171 (1999) ("There are good reasons for making this assumption in the typical civil case. . . . There will usually be no reason for valuing the defendant's rights more than the plaintiff's rights; consequently, there is no reason for preferring an error in one direction to one in the other." (This view is contrasted with the criminal setting in his immediately preceding discussion.)).

8. *See, e.g.*, Addington v. Texas, 441 U.S. 418, 423 (1979) ("At one end of the spectrum is the typical civil case involving a monetary dispute between private parties. Since society has a minimal concern with the outcome of such private suits, plaintiff's burden of proof is a mere preponderance of the evidence."); GENE R. SHREVE & PETER RAVEN-HANSEN, UNDERSTANDING CIVIL PROCEDURE 409 (3d ed. 2002) ("The function of that level or standard of proof is to 'instruct the fact-finder concerning the degree of confidence our society thinks he [or they] should have in the correctness of factual conclusions for a particular type of adjudication.' In most civil cases society is only minimally concerned with outcome, and the level of certainty is usually expressed as a preponderance of the evidence . . . ." (quoting *In re* Winship, 397 U.S. at 370 (Harlan, J., concurring))).

deriving from there being no other focal points—besides 0% and 100%, neither of which has any appeal.[9]

For criminal cases, the reasoning is more readily identified, mention often being made of the high stakes that make false convictions particularly problematic.[10] Even here, elaboration is rare and the conclusions are not obvious: High stakes make erroneous acquittals more troublesome as well; note that multiplying the consequences on both sides of a balance by a common factor has no effect on which way the scale tips.[11] Moreover, as will be elaborated in the body of this Article, stricter proof burdens can, in plausible settings, increase rather than decrease the number of false convictions, and the presence of higher social costs of sanctions likewise has ambiguous implications regarding whether the proof burden should be higher or lower.[12] The point of these observations is not that existing evidence thresholds are too high or too low in either the civil or criminal settings but rather that current thinking—actually, fairly old thinking that has been repeated but not much reconsidered—provides an insufficient basis for addressing the question.[13]

9. *See, e.g.*, FINKELSTEIN, *supra* note 3, at 66 ("A higher standard would cast a disproportionate burden on the proponents and would be arbitrary, because there would seem no rationale for selecting any particular decision probability above 0.50.").

10. *See, e.g.*, *In re* Winship, 397 U.S. at 372 (Harlan, J., concurring) ("In a criminal case, on the other hand, we do not view the social disutility of convicting an innocent man as equivalent to the disutility of acquitting someone who is guilty."); *Addington*, 441 U.S. at 423 ("In a criminal case, on the other hand, the interests of the defendant are of such magnitude that historically and without any explicit constitutional requirement they have been protected by standards of proof designed to exclude as nearly as possible the likelihood of an erroneous judgment."); 1 WAYNE R. LAFAVE, SUBSTANTIVE CRIMINAL LAW 77 (2d ed. 2003) ("It is a basic policy of Anglo-American criminal law that, in view of the serious consequences which follow conviction of crime, the prosecution has the burden of proving beyond a reasonable doubt all the facts necessary to establish the defendant's guilt."); 2 MCCORMICK ON EVIDENCE, *supra* note 3, at 491 ("However, this is not the case in a criminal action. Society has judged that it is significantly worse for an innocent person to be found guilty of a crime than for a guilty person to go free. The consequences . . . are usually more serious than the effects of an erroneous judgment in a civil case.").

11. *See infra* note 55. For some critical views in this regard within the criminal context, including by Adam Smith and Pierre-Simon Laplace, see note 42. *See also* DAVID CRUMP ET AL., CRIMINAL LAW: CASES, STATUTES, AND LAWYERING STRATEGIES 265-66 (2d ed. 2010) (raising but not answering questions about whether the standard ethical balance in the criminal setting is so clearly correct).

12. *See infra* Subsection II.C.1 and Section IV.B (on how the height of the evidence threshold affects the frequency of mistaken imposition of liability); Subsection III.B.2 (on how socially costly sanctions affect the optimal evidence threshold).

13. *Cf.* MORGAN, *supra* note 5, at 86 ("The truth is that in allocating the burden of persuasion and explaining to the jury their duties with reference to it, the courts have been performing their functions with a minimum of efficiency and with what President Eliot of Harvard once

To further motivate the inquiry, consider the application of conventional conceptions of the burden of proof in adjudication to the context of medical decisionmaking.[14] Would we say that, because society per se—as distinct from the patient—has no special stake in most medical treatment decisions, the rule for whether to have surgery, receive chemotherapy, or be subject to some other procedure should be based on a fifty percent probability? Or that, if the stakes are high (say, intrusive surgery for a very serious medical condition), it is better to let ten diseased patients go untreated than to mistakenly undertake surgery on one who did not need it? Or, perhaps the opposite, that it is better to let ten nondiseased patients suffer side effects than to let one diseased individual go untreated? Such formulations are confused, indeed absurd. Surely in setting the treatment threshold we should weigh, along with the likelihood that the patient truly has the condition, the expected benefit of the treatment to those who need it (would they otherwise die? what are the side effects and other costs? are there alternatives?) and the expected cost to those who in fact do not. And if we were observing doctors and patients making medical treatment decisions, we would expect to see them balancing the consequences of the possible outcomes—treatment and abstention—for the truly sick and for those who only appear to be ill, factoring in the likelihoods of each.

Legal systems, like medical systems, are of great importance. Civil and criminal law underlie social order, playing an essential role in facilitating economic activity, ensuring public safety, and otherwise promoting social welfare. Accordingly, decision criteria for adjudication—the setting of proof burdens in various legal contexts—should rest on a stronger foundation than age-old dicta. They should instead be grounded in explicit analysis that attends to the consequences of legal outcomes: correct and mistaken imposition of liability as well as proper and erroneous exoneration.

The pertinent effects are familiar in rough terms even though they have not systematically informed discourse on how proof burdens ought to be set. For individuals who might commit harmful acts, the prospect of liability produces deterrence whereas the possibility of mistaken absolution dilutes it. Deterrence of wrongful conduct is central not only to controlling crime but also to inducing individuals to perform contracts, comply with environmental and

described as a maximum of intellectual frugality. It is high time that the entire subject be critically reexamined.").

14. As elaborated in Section IV.C, the analogy is fairly close regarding legal rules that regulate future conduct, but there are substantial differences when legal rules influence ex ante behavior. Despite the latter divergence, the example aptly illustrates some of the shortcomings of much existing understanding.

health regulations, avoid careless activity, and interact honestly in the marketplace. The control of harmful behavior is, of course, the raison d'être for the legal system, and it is crucial to consider how the system's ability to achieve this objective is affected by how high the burden of proof is set.

The strength of the burden of proof also determines the magnitude of the legal system's major negative consequence aside from direct operational costs: namely, erroneous assignments of liability. A nontrivial likelihood that those who commit benign acts might be found liable will tend to chill desirable behavior. Although chilling receives less explicit attention in general treatments of the legal system and in discussions of the burden of proof, it is nevertheless a significant consideration in most areas of substantive law. Just as the anticipation of proper assignment of liability encourages mutually beneficial contractual interaction, concerns for mistakes may discourage it. Misapplication of antitrust law may chill aggressive competition or useful collaboration; with securities law, costs of raising capital may be inefficiently elevated; with medical malpractice, doctors may avoid high-risk patients or treatments and otherwise engage in defensive medicine.[15]

Part I of this Article investigates in a basic setting how to set the evidence threshold optimally in light of deterrence and chilling effects.[16] Slightly

15. For empirical evidence suggesting the importance of these phenomena, see, for example, Thomas J. Walker et al., Legal Opportunism, Litigation Risk, and IPO Underpricing (July 2010) (unpublished manuscript), http://www-rohan.sdsu.edu/~kpukthua/publications/JLE-6367.pdf; David M. Studdert et al., *Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment*, 293 JAMA 2609 (2005); and Daniel P. Kessler, William M. Sage & David J. Becker, *Impact of Malpractice Reforms on the Supply of Physician Services*, 293 JAMA 2618 (2005).

16. That is, deterrence and chilling are analyzed in a stripped-down setting that focuses on essential features while abstracting from various complicating factors, many of which are taken up later in the Article. The analysis differs from prior law and economics work on the burden of proof, or on errors, because other investigations focused on ex ante behavior usually contemplate that individuals' only behavioral decision is whether to commit a single harmful act—or, equivalently, whether to take precautions, where commission of the act is taken as given. *See* sources cited *infra* note 102. The results in such models are radically different from those derived here when there is a concern both for deterrence and for the chilling of other, desirable behavior. For a formal demonstration of the contrast, see Louis Kaplow, *Optimal Proof Burdens, Deterrence, and the Chilling of Desirable Behavior*, 101 AM. ECON. REV. PAPERS & PROC. 277 (2011) [hereinafter Kaplow, *Proof Burdens*], and for a more comprehensive formal analysis of the subject, see Louis Kaplow, On the Optimal Burden of Proof (Jan. 26, 2011) (unpublished manuscript) (on file with author) [hereinafter Kaplow, Optimal Burden of Proof]. *See also* Ivan P.L. Png, *Optimal Subsidies and Damages in the Presence of Judicial Error*, 6 INT'L REV. L. & ECON. 101 (1986) (in investigating errors, considering both the frequency with which individuals engage in an activity and their level of care, but not analyzing the burden of proof).

reducing the evidence threshold will increase the probability that individuals who contemplate the commission of harmful acts would expect to be subject to sanctions, and likewise for those who contemplate benign acts.[17] The extent of each effect depends on the nature of the evidence in a manner that is described. It also is necessary to determine how many harmful and benign acts are deterred and chilled, respectively, on account of these increases in expected sanctions. This magnitude is shown to depend on how many individuals have private benefits for the two types of acts that approximately equal the corresponding expected sanctions. Finally, for given numbers deterred and chilled, the effects on social welfare need to be assessed. In both cases, when acts are discouraged, individuals' private benefits from these acts are forgone; in the case of harmful acts, society is spared their negative consequences. The calculus for determining the optimal evidence threshold is, on reflection, conceptually straightforward and in accord with welfare-based intuitive reasoning. As a practical matter, however, the factors are many, their magnitudes undoubtedly vary greatly across contexts, and ascertaining the pertinent quantities is likely to be difficult.

Part II juxtaposes this welfare-based analysis with conventional thinking about the burden of proof, taking standard formulations on their face, in idealized form, until the closing Section. To begin, it offers a precise expression for the preponderance rule in the present setting—noting that this formulation requires only slight modification to characterize other proof requirements, such as one requiring a seventy-five percent or ninety-five percent likelihood as a prerequisite for liability. Next is the comparison. The central finding is that there is almost no overlap between the direct determinants of the preponderance rule (or other such rules, including proof beyond a reasonable doubt) and those for the optimal evidence threshold.[18]

17. The analysis for raising the threshold is the same—with all the effects reversed—so it is not repeated. Generally, if a slight reduction in the threshold has net negative welfare effects, then some increase would have positive effects.

18. Another interesting finding is that, even though the preponderance rule and other conventional ones have a seemingly simpler algebraic formulation than that for the welfare-maximizing rule, the information requirements to apply them in a given case—which are daunting—are similar to those needed to ascertain the optimal rule. (This claim may seem surprising in light of the statement in the text that the direct determinants differ. The explanation is that essentially the same underlying information is utilized quite differently under the two approaches in devising an evidence threshold.) Of course, even if the preponderance rule and others were easier to employ, they could hardly be commended on that account because their facility would be due to their ignoring the welfare-relevant effects.

An immediate implication of this divergence is that, depending on the context, the optimal evidence threshold could be much more demanding or notably more lax than the preponderance rule (or a requirement of proof beyond a reasonable doubt).[19] For example, if much benign activity might be chilled and the harmful acts in question cause little social loss and are mostly undeterrable in any event, very strong evidence should be required for liability. In contrast, if there is little prospect of chilling beneficial activity and the pertinent harmful acts impose extreme damage and might readily be deterred, a low threshold should be employed.

On reflection, this disjunction is not surprising because the many elements that determine the magnitudes of deterrence and chilling effects do not pertain to whether it is more likely than not that the individual before the tribunal committed a harmful act rather than a benign one—and conversely. Two central differences explain this gulf. First, the preponderance rule, and other rules based on the likelihood that the individual before the tribunal is one who committed the harmful act, embody an ex post perspective that takes behavior as given (exogenous) and asks, in light of that behavior, what is the likelihood of one versus another characterization. By contrast, a welfare-based, optimal threshold in these settings is determined by asking how behavior (taken to be endogenous) will change as a function of a change in the evidence threshold. Second, as highlighted by the medical treatment example, welfare-based thresholds depend on the harms and benefits of outcomes whereas likelihood-based formulations do not.

Part II continues by exploring other features of the preponderance rule (and other qualitatively similar rules). One striking, unappreciated point is that imposing a higher evidence threshold—requiring stronger evidence as a prerequisite for liability—could reduce the likelihood that, when evidence is just at the new, higher threshold, the case at hand involves one in which the individual in fact committed a harmful act. Accordingly, it is possible, starting from an evidence threshold that implements the preponderance rule, that substitution of a higher, more demanding threshold could have the property that cases with evidence barely above that new threshold fail the preponderance test. The reason is that raising the evidence threshold reduces the chilling effect, so more benign acts are committed, more will tend to come

19. Even on the high end, note that demanding proof beyond a reasonable doubt does not in practice require absolute certainty. If it was understood to require, for example, a 95% likelihood, it may well be that in many cases this demand could readily be met, although insisting on a 99% or perhaps a 99.99% likelihood would be prohibitive in most instances. *See also* sources cited *infra* note 78 (showing a wide range of probabilities associated with the requirement of proof beyond a reasonable doubt, most notably below 100%).

before tribunals, and therefore more close cases may involve benign rather than harmful acts despite the higher threshold. This effect is one of many and will hardly dominate in all situations, but it is entirely plausible that it will in some. Closely related, demanding stronger proof could increase the number of mistaken findings of liability.

In addition, Part II examines other proffered criteria for the burden of proof, such as seeking a high ratio of correct to incorrect impositions of liability, a minimal number of mistaken findings of liability, or an equal likelihood of mistakes for those committing harmful acts and benign ones. These and other, related desiderata are also shown to have virtually no connection to the optimal evidence threshold. As well, they have anomalous implications of their own. For example, a system with perfect deterrence and rare mistakes will have a zero ratio of correct to mistaken findings of liability, whereas a system in which all engage in harmful acts and liability is random will have an infinite ratio of correct to mistaken findings of liability. In addition, bringing additional individuals who did not commit harmful acts into the system can readily improve the ratio of correct to incorrect outcomes even if some of the additional cases result in mistaken findings of liability. And attempts to equalize the likelihoods of mistakes of the two types could require an intentional increase in imposing liability on the wrong individuals if it is difficult to reduce a fairly substantial rate of mistaken exoneration.

In sum, Part II indicates that conventional thinking about proof burdens is deeply flawed, both in terms of its own internal workings and regarding its (lack of) relationship to promoting social welfare. This Part concludes by inquiring whether conventional conceptions—in their idealized form, as officially pronounced by courts and analyzed by commentators—are fully reflected in current practice or perhaps instead are applied in ways that implicitly embody some consequentialist elements of the sort developed in Part I. Relatedly, it offers some thought experiments about alternative formulations of the burden of proof that entail more explicit attention to the consequences of adjudication for social welfare.

The remainder of the Article returns to its core conceptual mission, the explication of optimal (welfare-promoting) policy, entertaining a wider range of factors and settings. Part III addresses other features of the legal system. In an important sense, the proof burden can usefully be viewed as one of the many enforcement instruments that a legal regime deploys to control harmful conduct. Section III.A examines the interaction between setting the level of enforcement effort, taking into account its costliness, and setting the evidence threshold. Section III.B analyzes the level of sanctions. In addition, it considers how the social costs of sanctions, notably imprisonment, relate to how the evidence threshold should be set. It turns out that higher social sanction costs

do not unambiguously favor a higher evidence threshold: even though a higher threshold reduces the rate of imposition of sanctions for the cases that come before tribunals, the resulting reductions in deterrence and chilling increase the flow of cases into adjudication, which has a countervailing effect on the frequency of use of costly sanctions. Likewise, if society has a greater concern for mistakenly imposing sanctions on the innocent, a higher evidence threshold can sometimes be counterproductive because innocents will more often come before tribunals—even though the sanction probability per innocent individual before a tribunal is reduced. Finally, Section III.C analyzes the interaction between proof burdens and the accuracy of the legal system: roughly speaking, the latter determines the overall error rate whereas the former dictates how to divide the errors between the two types. Different levels of accuracy may bear on the optimal evidence threshold, and vice versa.

Part IV introduces additional methods of enforcement. Until this point, the analysis focuses (in a manner detailed in Section I.A) on a particular enforcement setting, one akin to monitoring—such as the posting of police or security guards—or auditing, including the use of inspections and spot-checks. Section IV.A elaborates on these methods and their differences, with particular attention to the administrative costs of enforcement. Section IV.B explores the technique of investigation, which refers to waiting until harm is observed (such as when there is a murder, stolen automobile, or pollution discharge) and then attempting to identify the perpetrator. The determinants of the optimal evidence threshold are different in important ways: there are more factors and their interaction is somewhat more complicated. Furthermore, there is an even greater contrast between this formulation and conventional ones—the preponderance rule and others based on the likelihood that the individual before the tribunal is a wrongful actor. Another finding is that it is plausible that a higher evidence threshold might actually increase how often sanctions are mistakenly imposed on innocent actors. The reason is that a higher threshold reduces deterrence, so more harmful acts are committed, which triggers more investigations, which in turn are the genesis of potential erroneous findings of liability.

Section IV.C considers an enforcement context that is qualitatively different from those analyzed previously: the regulation of future conduct, such as through allowing zoning variances, granting licenses, authorizing drugs, or permitting mergers. Unlike with contract law, liability for securities fraud, and most tort and criminal law, in these settings the central concern often is not with prior conduct—which can be influenced (through deterrence and chilling) by the prospect of sanctions—but with future behavior. The decision at hand is whether to allow or prohibit some action, going forward. Here, the determination of the optimal evidence threshold differs quite substantially

from that in the prior settings. Instead, it is akin to the sort of analysis used in making medical treatment decisions, as described at the outset of this Introduction. This formulation, in turn, is notably closer to the preponderance rule (and other such likelihood-based rules), although it still differs markedly by the explicit inclusion of the harms and benefits that may flow from different outcomes. For example, in deciding whether to approve a drug application, the government agency will not simply ask, say, whether the drug is more than fifty percent likely to be beneficial rather than harmful; it will also place great weight on the magnitudes of the benefits and harms. Benefits may accrue to fewer than fifty percent of patients, but if the benefits are dramatic, the drug will be approved even in the presence of certain adverse side-effects, if they are sufficiently mild. Interestingly, in this context in which the optimal rule is less qualitatively distinct from the preponderance rule, one does not ordinarily see the preponderance rule employed, presumably because the residual differences are obvious—and considerable.

Section IV.D briefly examines some aspects of how public and private modes of enforcement determine the cases that enter the legal system.[20] Emphasis is placed on the motives of public enforcers and private litigants, which unfortunately may diverge significantly from advancement of the social good. These incentives are of particular relevance here because of possible interactions with the burden of proof. For example, demanding stronger evidence may make it harder for police and prosecutors to abuse their authority in pursuit of a regime's political opponents. This effect is one of the familiar justifications for requiring proof beyond a reasonable doubt in criminal cases, although as some have noted the rationale can have similar force in civil settings in which the government is the enforcer.

This Article offers a preliminary, conceptual analysis of the optimal stringency of the burden of proof. Prior literature has devoted surprisingly little attention to this question. Conventional notions that focus on the likelihood that the individual before the tribunal has committed a harmful act turn out to be almost entirely unrelated to welfare-based analysis and also to pose some unrecognized internal conundrums. The present analysis breaks new ground across much of the territory—in terms of enforcement settings, legal system features, and pertinent factors—but it is hardly comprehensive or definitive along any of these dimensions. Moreover, in order to undertake the

20. This Part comments briefly on the related question of parties' incentives to develop evidence. Incentives to sue and to present evidence, and attendant costs, have been a focus of much prior law and economics writing on the burden of proof (often addressing the placement of the production burden). Some of this literature is cited in note 5 and in Section IV.D.

important constructive task of making sensible recommendations for system design, one would need context-specific empirical evidence that is not readily available. Nevertheless, it is difficult to collect the requisite evidence or even to make reasonable conjectures until one has identified what factors are relevant, why, and how. The inquiry must begin somewhere if it is to begin at all.

I. OPTIMAL EVIDENCE THRESHOLD

A. Framework

1. Context

Before discussing the nature of evidence and the evidence threshold that must be crossed for liability to attach, our central interest, it is important to describe the context in which adjudication is understood to arise. To begin, suppose that some individuals have opportunities to commit harmful acts, the prospect of which is the principal justification for the legal system. A harmful act is understood here to be one that generates some level of external harm—that is, to a third party—in addition to whatever benefit accrues to the individual who commits it.[21] Individuals' benefits from such acts are assumed to vary. Although it is not ruled out that some individuals' benefits may exceed the harm caused, in which case the so-called harmful acts would be net socially desirable, attention will largely be confined to settings in which the level of the expected sanction is such that acts that are deterred involve a net social harm.[22]

In the absence of any legal restraint, it is assumed that individuals will act whenever their private benefit from doing so is positive. With a legal system, they are assumed to act if and only if their private benefit exceeds the expected

21. For ease of exposition, each harmful act is imagined to cause only one level of harm and to cause this harm with certainty. Allowing for variable or uncertain harm (in which case expected harm would replace the level of harm) would not materially affect the analysis. Acts that cause different levels of expected harm may be viewed as different types of acts that may, in principle, be subject to different rules (for example, different sanctions and different evidence thresholds). An important and more subtle variation, not analyzed here, is the case where there is unobservable variation in the level of harm that is correlated with the strength of the evidence likely to be generated; the main effect on the analysis would be that the level of harm, for example, in the lower element of the left box in Figure 4 in Section I.B would refer to the level of harm conditional on the evidence threshold under consideration.

22. If the marginal deterred harmful act is net socially desirable, it will be obvious (in basic settings) that the legal system is not designed optimally. Social welfare would be greater with, for example, lower sanctions, lower enforcement effort, or a higher evidence threshold. Most of the tradeoffs elaborated throughout this Article would not arise.

sanction.[23] Accordingly, it is necessary to describe how the legal system generates sanctions.

Enforcement is taken to operate in two stages. First, some sort of scrutiny brings a portion of individuals into the legal system on account of the acts that they are observed to commit. Second, for those brought into the system, adjudication assesses the evidence in order to decide whether to exonerate or instead to assign liability and impose the applicable sanction—taken here to involve a monetary payment, which might be a fine paid to the state or damages paid to a plaintiff.[24] (Nonmonetary sanctions are considered in Subsection III.B.2 and outcomes that involve the regulation of future conduct are analyzed in Section IV.C.) If it were known that all individuals brought into the legal system in the first stage indeed committed harmful acts, adjudication would be simple; accordingly, we now consider how individuals who have not committed harmful acts sometimes enter the system.

Another group of individuals is assumed to have opportunities to commit benign acts. A benign act (like a harmful act) generates a benefit to the individual who commits it, but (unlike a harmful act) it does not generate any external harm. Individuals commit benign acts because of the benefits they receive, and since there is no external harm, these acts are net socially beneficial. (Note that the labels "harmful" and "benign" refer to the presence or absence of external social harm, not to the act as a whole.) For ease of exposition, assume that the two groups of individuals are distinct; that is, one only has opportunities to commit harmful acts and the other only to commit benign acts. The analysis would be much the same if given individuals sometimes had opportunities to commit one type of act and, at other times, opportunities to commit the other. Also, the analysis would change somewhat

23. A number of simplifying assumptions are implicit. First, individuals are assumed to act to maximize their self-interest. One could introduce other influences, such as an internal desire to abide by the law or the presence of social sanctions, including reputational loss, without greatly affecting the analysis. (As a first cut in analyzing such a case, we could reduce the valuation of the private benefit from acts subject to such costs accordingly; note that in some cases, particularly regarding harmful acts, this adjustment might convert an otherwise privately beneficial act into one involving a net negative private benefit, in which case the act would not be committed even in the absence of legal sanctions.) Second, individuals are assumed to be risk neutral. Risk aversion also would not qualitatively affect much of the analysis. *See infra* note 149; *see also infra* note 199 (discussing the compensatory function of private litigation, which is set to the side here).

24. The present discussion abstracts from settlement (including plea bargaining), a possibility considered briefly in Section IV.D. It may be supposed that settlements (and pleas) involve sanctions equal to the expected value of sanctions in adjudication.

but not drastically if individuals at the same time could choose whether to engage in the harmful act, the benign act, or no act at all.[25]

It is assumed that the enforcement apparatus in the first stage brings some individuals who commit benign acts into the legal system. They too enter adjudication, so the adjudicator's task will be the familiar one of trying to distinguish the two situations and, to the extent feasible, confine sanctions to those who committed harmful acts, without exonerating very many of those individuals.

For both groups of individuals, the prospect of sanctions will influence decisions whether to commit acts in the first instance. An individual with an opportunity to commit a harmful act will be *deterred* from doing so when the private benefit of the act is less than the expected sanction, which is the product of three factors: (1) the likelihood that harmful acts will be identified in stage one, (2) the likelihood that an individual in adjudication at stage two who committed a harmful act will be subject to sanctions, and (3) the magnitude of the sanction. Likewise, an individual with an opportunity to commit a benign act will be *chilled* from doing so when the private benefit of the act is less than the expected sanction, which also will be the product of three corresponding factors. Although deterrence and chilling are analytically the same, different terms help keep the phenomena distinct to reflect their differing social consequences: deterrence is desirable whereas chilling constitutes an undesirable side-effect of the legal system.

Although the three factors are conceptually the same for each type of act, the levels of the first two generally will differ. (The magnitude of the sanction does not because, at the time of its application, the legal system will have failed to distinguish the two types of individuals.) The second factor, involving evidence assessment and the assignment of liability, is considered in detail in the next Subsection. The first, concerning the likelihood that the two types of acts are identified at the first stage of scrutiny, will depend on the method of enforcement. For Parts I through III, attention will largely be confined to methods of enforcement that are akin to monitoring (for example, officers

25. The main difference in this final scenario is that, when deterring a marginal harmful act, there is the chance that an individual would switch to a benign act instead of inaction; because the benign act generates a benefit, the net social value of deterrence would be greater on this account. Likewise, when chilling a marginal benign act, some individuals might switch to a harmful act, which would make the net social cost of chilling greater. Accordingly, the stakes regarding both behavioral effects would be raised, but much of the analysis, often concerning tradeoffs between the two types of behavioral effects, would be qualitatively similar. (Another generalization, subject to similar analysis, would be to allow different choices to cause different, positive levels of harm – for example, the choice among a more harmful act, a less harmful act, and inaction.)

posted along a highway or on the beat) or auditing (including random inspections, such as for environmental or fire code violations). Specifically, some fraction of each type of act is identified (typically, we would hope, a larger portion for the harmful type of act), and those fractions are determined by the intensity of enforcement effort (taken for now to be fixed).[26] Such matters will be reexamined in some detail in Part IV, which considers alternative methods of enforcement and how their differences are consequential for setting the evidence threshold. Also, for the present (until Section III.A), the costs of enforcement will be set aside.[27]

As mentioned in the Introduction, the central problem arises because certain sorts of benign acts are sometimes confused with harmful ones. Therefore, it is useful to think of the benign acts under consideration as limited to that type. For speeding, we would restrict attention to individuals who in fact are driving vehicles and perhaps only those driving near the speed limit rather than substantially under it. For securities laws regulating disclosures at public offerings, we would consider as pertinent benign acts only other public offerings, specifically, ones accompanied by valid disclosures but in circumstances where it might be hard to tell, readily and for sure, whether appropriate disclosures were made.[28] These are the sorts of benign acts that

26. The analysis here supposes that whether one or another harmful or benign act is swept into the system is independent of how many of the two types of acts occur or how others are scrutinized at this stage. (As we will see in Section IV.B, this assumption does not hold for enforcement by investigation, wherein scrutiny is triggered by the observation of a harmful act.) This assumption would not hold, for example, if enforcers' incentives created interdependence, such as when a police officer or inspector feels pressure to generate a certain amount of fine revenue, in which case greater deterrence of harmful acts may induce enforcers to identify benign acts as harmful ones. Public enforcers' incentives are considered in Subsection IV.D.1.

27. To the extent that enforcement effort is being held constant, as is supposed, this assumption is not problematic for the pure case of monitoring at stage one (and with a cost-free stage two). More generally, however, as Section IV.A explores, enforcement costs are influenced by deterrence and chilling, which in turn are affected by the evidence threshold, so a complete analysis cannot ignore these costs.

28. If one defined the set of benign acts more broadly, little would change because the rates of identification at stage one and assessment of sanctions after adjudication would fall. Importantly, however, the expected sanction facing individuals contemplating benign acts would vary in that case. Those contemplating acts that would rarely if ever be confused with harmful ones would face an expected sanction essentially equal to zero, whereas those contemplating acts that could plausibly be so confused would face a nontrivial expected sanction and hence may be chilled. Realistically, the likelihood of confusion (in those cases where such is conceivable) will vary across benign acts, and thus the intensity of chilling effects will differ as well.

might be brought into the legal system in stage one, and hence they are the ones that adjudicators must distinguish at stage two, to which we now turn.

2. Evidence Threshold

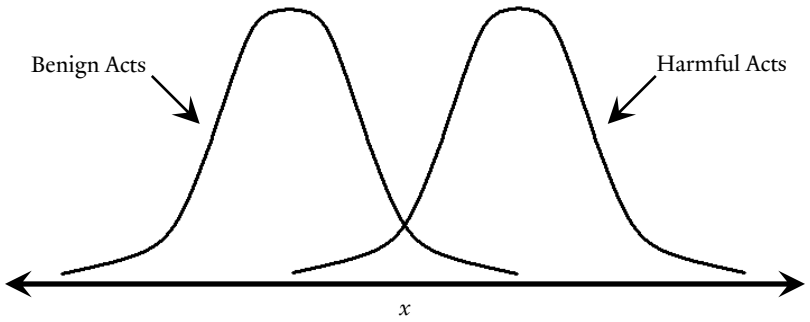
To introduce evidence and the setting of an evidence threshold, consider a simple and reasonably familiar example in the medical context: the use of test results to determine whether to treat a patient—perhaps to administer a drug, perform surgery, or employ a more invasive follow-up diagnostic procedure. The test is an imperfect signal, higher scores indicating a greater likelihood that the disease in question is present. That is, individuals who truly have the ailment produce a range of test results, but their scores cluster toward the high end, whereas individuals who really are disease-free also produce a range of results, but their scores cluster toward the low end. The problem is to choose a cutoff or threshold, above which treatment will be applied. A high cutoff will result in few false positives, which is to say that only a small portion of disease-free individuals will be mistakenly given the treatment; however, a high cutoff will also result in many false negatives, so a nontrivial fraction of diseased individuals will mistakenly fail to receive treatment. In determining the optimal threshold, these error costs will be traded off: if nontreatment of diseased individuals is serious and the treatment involves little cost to disease-free individuals, a low cutoff will be optimal, but if nontreatment is only moderately problematic whereas treatment is very costly to disease-free individuals, a high cutoff will be optimal. (It turns out that the optimal decision rule for this case is much like that considered in Section IV.C, on the regulation of future conduct, but substantially different from that appropriate in our present context. Nevertheless, the example is useful here because it does properly depict the nature of evidence and of evidence thresholds, the present subject.)

Returning to our legal context, it is supposed that identification of an individual at the first stage in the law enforcement process will give rise to a body of evidence that the tribunal must assess.[29] In many instances, the evidence may be multidimensional and murky, unlike the simple medical test just described. Of course, actual medical decisionmaking typically involves many sorts of evidence, some of which are much less sharp than a blood test result. For present purposes, we will collapse all the evidence into a single indicator, x, of the overall strength of the evidence because our focus is on how

29. How such evidence is generated will not be addressed, except in passing in Parts III and IV.

strong a case should be required for liability.[30] In accord with the medical diagnosis example, harmful and benign acts each generate a wide range of possible values of x,[31] but values for harmful acts will tend to cluster at a higher level than do values for benign acts, as depicted in Figure 1.

Figure 1.
PROBABILITY DISTRIBUTIONS FOR STRENGTH OF EVIDENCE



These curves are drawn roughly as normal distributions with different means and the same variance merely for the sake of illustration.[32] Also, no particular

30. There is a relationship between defining the strength of evidence and combining multiple pieces of evidence. Formally, if one wishes to minimize the likelihood of one type of error while holding the rate of the other type of error constant, it is optimal to order evidence by the likelihood ratio: the ratio of the likelihood that such evidence would be generated by a harmful act to the likelihood that such evidence would be generated by a benign act, where the likelihoods correspond to the height of the probability distributions in Figure 1. This concept has multidimensional analogues, and the general principle is given by the Neyman-Pearson Lemma. *See* J. Neyman & E.S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, 231 PHIL. TRANSACTIONS OF THE ROYAL SOC'Y OF LONDON, SERIES A, CONTAINING PAPERS OF A MATHEMATICAL OR PHYSICAL CHARACTER 289 (1933); *see also* NEIL A. MACMILLAN & C. DOUGLAS CREELMAN, DETECTION THEORY: A USER'S GUIDE 141-63 (2d ed. 2005) (explaining the extension of signal detection theory to multiple dimensions).

31. It is assumed that, ex ante, all acts of a given type generate the same probability distribution of evidence. Allowing for heterogeneity in this regard (that is, for a given type of act) would complicate the analysis without altering its qualitative features. Likewise, one could allow for the nature of the enforcement process—*see infra* Section III.A and Part IV—to influence these distributions, conditional on cases appearing before a tribunal. See especially the discussion at the end of Subsection IV.D.2.

32. This special case is typically used in presenting basic signal detection theory, which is the first formally developed system that closely corresponds to the construction used here. *See generally* MACMILLAN & CREELMAN, *supra* note 30; R. Duncan Luce, *Detection and Recognition*, *in* 1 HANDBOOK OF MATHEMATICAL PSYCHOLOGY 103 (R. Duncan Luce, Robert R. Bush & Eugene Galanter eds., 1963). This theory was originally developed for radar

scale for x or zero point is indicated because, for present purposes, all that matters is the ordering, namely that higher values of x are taken to indicate stronger evidence.[33]

Given this depiction of evidence, the legal system must choose an *evidence threshold*, denoted here by x^T, which indicates the value of x above which liability will be assigned and below which there is no liability. It is likewise conventional to suppose that ties go to the defendant, so that, in cases in which x equals x^T, there is no liability.[34]

Now, for any given evidence threshold, x^T, we can use Figure 1 to determine the probability that (conditional on coming before the tribunal) an individual who has committed a harmful act will be found liable. This will be denoted $P_{HARMFUL}(x^T)$. We can also determine the probability that one who has committed a benign act will (mistakenly) be found liable: $P_{BENIGN}(x^T)$. Suppose, for example, that x^T is set at the value of x in the middle of the figure, where the two curves happen to cross. In that case, $P_{HARMFUL}(x^T)$ will be given by the portion of the area under the Harmful Acts curve that lies to the right of this value of x^T (which looks to be roughly 95% of the area) because, for all such higher values of x, the individual would be found liable (and correctly so). Similarly, for all lower values, the individual is (mistakenly) exonerated, and

detection. *See, e.g.*, J.I. Marcum, *A Statistical Theory of Target Detection by Pulsed Radar* (U.S. Air Force Project Rand, Research Memorandum No. 754, 1947), *available at* http://www.rand.org/content/dam/rand/pubs/research_memoranda/2006/RM754.pdf.

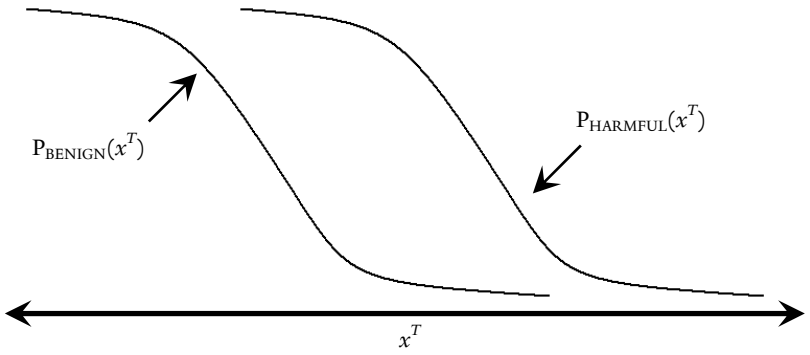
33. Interestingly, Bentham, in the course of advocating that factfinders not merely count witnesses but instead consider the strength of the testimony, suggested that each witness be assigned a value on a scale from -10 to 10, where 0 was a neutral point, negative values favored nonexistence of a fact, and positive values denoted existence. One would then sum such values across witnesses and find a fact to exist if and only if the sum was positive. *See* 1 JEREMY BENTHAM, RATIONALE OF JUDICIAL EVIDENCE, SPECIALLY APPLIED TO ENGLISH PRACTICE 71-109 (Rothman & Co. 1995) (1827).

34. Strictly speaking, with continuous probability distributions as in Figure 1, the probability that x will precisely equal any given x^T is zero, making ties extremely unimportant. (Readers not familiar with this feature of probability theory may ponder the likelihood that, say, it will be precisely twelve noon when one looks at one's watch, as distinguished from 11:59:59.999999999 or 12:00:00.000000001, and so forth.) It is familiar, by contrast, that the legal system—as reflected in legal commentary, lawyers' arguments, and court opinions—attaches great weight to tiebreaking, which suggests that something else is going on. One phenomenon, not considered in this Article, involves production burdens. *See supra* note 5. Another is that factfinders psychologically may perceive a nontrivial range of close cases as sufficiently difficult to decide to be viewed as ties. As a consequence, a more-likely-than-not requirement might, in many settings, result in factfinders perceiving a tie (and thus awarding a victory to the defendant) when the probability is less than, say, 55%. It may be helpful to think of such settings as involving a de facto requirement of 55% rather than 50%.

this probability, 1 - $P_{HARMFUL}(x^T)$, is given by the portion of the area under the Harmful Acts curve that lies to the left of this value of x^T (about 5%). Likewise, $P_{BENIGN}(x^T)$ will be given by the portion of the area under the Benign Acts curve that lies to the right of this value of x^T (approximately 5% of the area) because, for all such higher values of x, the individual would be found liable (by mistake). Similarly, for all lower values, the individual is (correctly) exonerated, and this probability, 1 - $P_{BENIGN}(x^T)$, is given by the portion of the area under the Benign Acts curve that lies to the left of this value of x^T (about 95%).[35]

We could instead set x^T higher, which would reduce $P_{BENIGN}(x^T)$, the likelihood in adjudication of false assignments of liability for benign acts, but this change would also reduce $P_{HARMFUL}(x^T)$, the likelihood of correct assignments of liability for harmful acts—which is to say, it would increase the frequency of mistaken exoneration. For lower values of x^T, the effects would be reversed. This relationship between x^T, on one hand, and $P_{BENIGN}(x^T)$ and $P_{HARMFUL}(x^T)$, on the other hand, is depicted in Figure 2.[36]

Figure 2.
PROBABILITIES OF LIABILITY AS A FUNCTION OF THE EVIDENCE THRESHOLD x^T



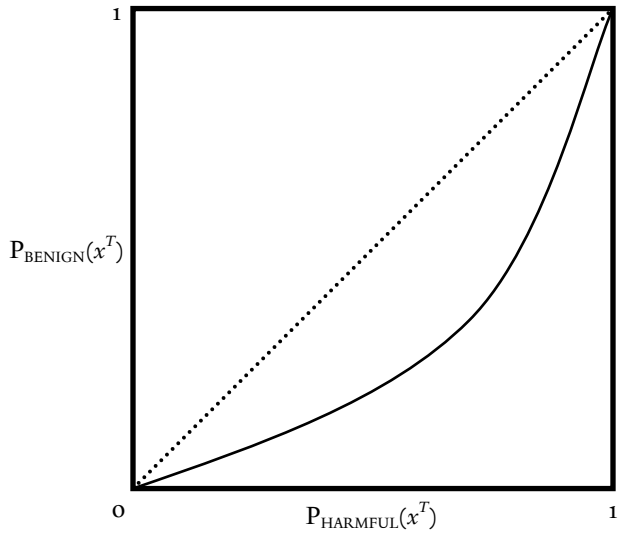
35. For the most part, language of correct or mistaken liability or exoneration will be employed in lieu of referring to mistakes as false positives and false negatives, or as type I and type II errors respectively, because in my experience individuals (including sophisticated, trained analysts) occasionally confuse these terms and, even more often, must hesitate to make the appropriate translations.

36. These curves, like those in Figure 1, are hand drawn rather than generated mathematically; hence their shapes bear only an approximate correspondence to those that would be implied by a formal translation of Figure 1. For readers familiar with probability theory, each of the probabilities in Figure 2, for a given value of x^T, is given by one minus the value of the cumulative distribution function corresponding to the probability densities in Figure 1. (The cumulative distribution function indicates the area to the left of x^T; hence, one minus this value gives the area to the right.)

As the evidence threshold x^T required for application of the sanction is raised, both probabilities fall. Initially (for a very low x^T), the probability of (mistakenly) sanctioning the benign act falls quite rapidly relative to the fall in the probability of (correctly) sanctioning the harmful act. However, as the threshold becomes fairly high, there is a sharper drop in the latter probability.[37]

The error tradeoff implicit in Figure 2 can be illustrated in another way. Suppose that the legal system wishes to set the evidence threshold x^T in order to achieve some target rate of liability for those before the tribunal who have committed harmful acts—that is, a target value of $P_{HARMFUL}(x^T)$. We can then ask what is the resulting rate of mistaken imposition of sanctions on benign acts—that is, $P_{BENIGN}(x^T)$. If one traces this relationship for different target values of $P_{HARMFUL}(x^T)$, the result would look something like that depicted in Figure 3.[38]

Figure 3.
$P_{BENIGN}(x^T)$ AS A FUNCTION OF $P_{HARMFUL}(x^T)$



37. Because there is no natural zero point on the horizontal axis, as previously mentioned, no vertical axis is drawn. However, it is supposed that the height of the two curves approaches one as we move sufficiently toward the left of the horizontal axis and both heights approach zero toward the right.

38. This sort of curve, like Figure 1, is also used in presentations of signal detection theory; there, the analogue to the curve in Figure 3 is called the receiver operating characteristic (ROC) curve, although it is conventional to reverse the axes, showing the false positive (called false alarm) rate on the horizontal axis and the true positive (called hit) rate on the vertical axis. *See, e.g.*, R. DUNCAN LUCE, INDIVIDUAL CHOICE BEHAVIOR: A THEORETICAL ANALYSIS 60-61 (1959); MACMILLAN & CREELMAN, *supra* note 30, at 9-11.

(For convenience, a 45° dotted line is shown, which indicates what the relationship would be if the evidence strength x had no discriminating power.[39]) Here, high values of the evidence threshold, x^T, would be toward the lower left of the figure, and low values toward the upper right: as stated in connection with Figure 2, high values correspond to low probabilities of liability, and low values give rise to high probabilities of liability. To restate this point to avoid possible confusion: raising the evidence threshold, x^T, means that one moves from the upper right corner toward the lower left corner—following the curve—and not the other way around.

This figure restates our tradeoff: the higher the value of $P_{HARMFUL}(x^T)$, the likelihood that adjudication will (correctly) assign liability when harmful acts are under consideration, the higher will be the resulting value of $P_{BENIGN}(x^T)$, the likelihood that benign acts will (mistakenly) be subject to sanctions. Moreover, given the depicted relationship, raising $P_{HARMFUL}(x^T)$ (i.e., lowering x^T) when $P_{HARMFUL}(x^T)$ is very low (x^T is very high) appears relatively advantageous in that $P_{HARMFUL}(x^T)$ rises rapidly relative to the concomitant increase in $P_{BENIGN}(x^T)$. But when $P_{HARMFUL}(x^T)$ is already high (x^T is low), we see that $P_{BENIGN}(x^T)$ rises rapidly relative to the concomitant increase in $P_{HARMFUL}(x^T)$.

The shape of the curve, rising at first slowly and then more rapidly, is implied by the shapes of the curves in Figure 2, which in turn derive from the curves in Figure 1.[40] In some fields, this relationship is described as the monotone likelihood ratio property.[41] This property captures the notion that stronger evidence is that which indicates a higher relative likelihood that an act is the harmful rather than the benign type. As mentioned, insisting on stronger evidence corresponds to moving along the curve toward the lower left corner of the box. And, given this curvature, as the evidence threshold gets more demanding, starting from a lax point, the probability that sanctions are applied to benign acts before the tribunal falls rapidly for a given decline in the probability of sanctioning harmful acts. By contrast, when the threshold is already fairly tough, the probability of sanctioning benign acts falls more

39. This would be so when the curves in Figure 1 are coincident (and likewise for Figure 2), which is to say that both types of acts are equally likely to result in any particular strength of evidence (value of x). In such a case, higher values of x would not be stronger in any meaningful sense.

40. *See supra* note 36.

41. *See, e.g.*, Samuel Karlin & Herman Rubin, *The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio*, 27 ANNALS MATHEMATICAL STAT. 272 (1956); Paul R. Milgrom, *Good News and Bad News: Representation Theorems and Applications*, 12 BELL J. ECON. 380 (1981).

slowly; after all, a rather low proportion of benign acts would still be subject to sanctions when the evidence threshold is already quite high.

B. Analysis

This Section analyzes how to determine the optimal evidence threshold in the setting presented in Section I.A. Optimality here refers to the extent to which a policy—the level of the evidence threshold—maximizes the aggregate welfare of individuals in society.[42] Put more concretely, the present objective is

---

42. Justice Harlan, in his concurrence to the Supreme Court's opinion in *In re Winship*, stated that "the choice of the standard for a particular variety of adjudication does . . . reflect a very fundamental assessment of the comparative social costs of erroneous factual determinations," and that "the choice of the standard to be applied in a particular kind of litigation should, in a rational world, reflect an assessment of the comparative social disutility of each." 397 U.S. 358, 370-71 (1970) (Harlan, J., concurring). He went on to claim that the U.S. legal system's standards in the civil and criminal contexts reflect such an assessment. *Id.* at 371-72. For a defense of a focus on individuals' welfare against alternatives that sometimes are advocated, particularly in the legal academy, see KAPLOW & SHAVELL, *supra* note 2 (especially Chapter V on legal procedure and Chapter VI on law enforcement). Nonwelfarist views are obviously influential in legal analysis, as indicated by many of the references cited in this book. Regarding proof burdens, see, for example, Redmayne, *supra* note 7, at 172 n.19, who finds that the effects of proof burdens on harm-causing accidents, safety costs, and prospective victim carelessness are irrelevant to his rights-based analysis.

Within a welfarist approach, there may also be concerns for distribution, which are not considered here. Note that, in civil settings, there is often substantial symmetry (consider rules of contract law and procedures for contractual adjudication, for example). And in criminal settings, some of the tradeoffs involve similarly situated individuals. For example, Laplace urged that one should treat on a par the prospect of harm to innocent individuals due to false convictions and the prospect of harm to innocent individuals due to a lack of incapacitation and a failure of deterrence. He formulated the relevant question as follows:

> Has the evidence for the offence with which the accused is charged the high degree of probability that is necessary in order that citizens may have less to fear from the errors of the courts, if the accused is innocent and convicted, than from his new attempts at crimes, and those of the unfortunate creatures who would be encouraged by the example of his impunity, if he is guilty and acquitted?

PIERRE-SIMON LAPLACE, PHILOSOPHICAL ESSAY ON PROBABILITIES 77-78 (Andrew I. Dale trans., Springer 1995) (5th ed. 1825); *see* ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 97 (1974) (offering as a central determinant of the proper standard of proof in criminal cases the following: "That system is most effective which minimizes the expected value of unearned harm to me, either through my being unjustly punished or through my being a victim of a crime."); Ronald J. Allen & Larry Laudan, *Deadly Dilemmas*, 41 TEX. TECH L. REV. 65 (2008) (advancing the view that mistaken exoneration is also a serious concern for innocent individuals and that it is inappropriate to confine attention to error rates at trials); *see also* ADAM SMITH, THE THEORY OF THE MORAL SENTIMENTS II.II.III.7 (D.D. Raphael & A.L. Macfie eds., Oxford Univ. Press 1976) (6th ed. 1790) (stating that "mercy to the guilty is cruelty to the innocent").

to maximize individuals' benefits from the acts they commit minus the harms that their acts cause. In our setting,[43] there are two effects of adjustments to the evidence threshold: effects on the deterrence of harmful acts and on the chilling of benign acts. Ideally, we would like to maximize deterrence[44] and minimize chilling, but in setting the evidence threshold, there is ordinarily a direct tradeoff, which will now be elaborated.

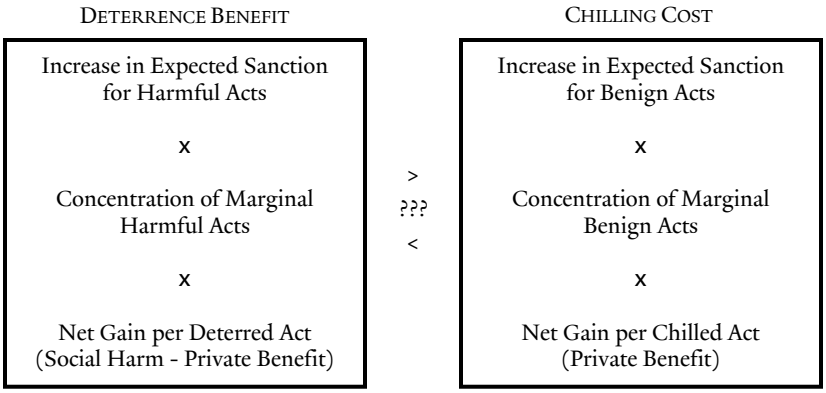
Suppose that the evidence threshold, x^T, is at some initial level and we contemplate lowering it slightly. Holding constant other features of the legal system—notably, enforcement effort and sanctions, and also taking the quality of available evidence as given—what will be the effects of this reduction? As is apparent from Subsection I.A.2, a lower threshold will increase $P_{HARMFUL}(x^T)$ and $P_{BENIGN}(x^T)$, the likelihood of applying sanctions to individuals before the tribunal who committed harmful and benign acts, respectively. Returning to the analysis in Subsection I.A.1, the consequence is to increase the expected sanctions for both types of acts. As a result, additional harmful acts will be deterred and more benign acts will be chilled. Accordingly, we wish to compare the aggregate deterrence benefit to the overall chilling cost; if the benefit is greater, the reduction is socially desirable, whereas if the cost is larger, the reduction is detrimental. Note that, in the latter case, it would tend to be beneficial to raise the evidence threshold somewhat, because the effects will be reversed. At the optimal evidence threshold, x^T, these competing forces will be in balance.

To undertake this comparison, we need to be able to quantify the two competing effects, which necessitates identifying their components. This decomposition is depicted in Figure 4.

43. Recall from Subsection I.A.1 that administrative costs and sanction costs are set to the side here, to be taken up in Parts III and IV.

44. More precisely, to set expected sanctions so as to deter all acts that are net socially harmful but not others, which is referred to as first-best deterrence.

Figure 4.
EFFECTS OF A REDUCTION IN THE EVIDENCE THRESHOLD



Let us begin by considering the left side of the figure. It indicates that the deterrence benefit of a reduction in the evidence threshold will equal the increase in the expected sanction for harmful acts weighted by the concentration of marginal harmful acts, which product is then multiplied by the net gain per act that is deterred—itself the difference between the social harm of the act and its private benefit. Each of these elements needs elaboration.

The product of the first two terms gives the number of individuals who will be deterred as a consequence of a reduction in the evidence threshold. Beginning at the top, we wish to know how much reducing the evidence threshold increases the expected sanction for harmful acts. Recall from Subsection I.A.1 that this expected sanction is the product of the likelihood that harmful acts are identified in stage one, the likelihood that an individual thus identified at stage one will be found liable in adjudication at stage two, and the magnitude of the sanction. Lowering the evidence threshold raises (only) the second component. But by how much?

This magnitude can be determined from either Figure 1 or Figure 2 in Subsection I.A.2. Using Figure 2, for a given initial level of the evidence threshold x^T, we can ask how much $P_{HARMFUL}(x^T)$ rises as x^T is reduced a bit, which is to say, as we move slightly to the left. This increment will be given by the (absolute value of the) slope of the $P_{HARMFUL}(x^T)$ curve. For example, if x^T is near the far left or far right of the figure, the increase would be small, but if x^T began somewhat right of center, where the curve is steeper, the increase would be larger.

The same information can be gleaned instead from Figure 1. The magnitude of the increase in $P_{HARMFUL}(x^T)$ is given by the height of the

probability distribution curve for the harmful act (which, again, is low at the far left and far right of the figure but high when right of center). As one moves the cutoff, x^T, slightly to the left, a small vertical slice of the area under the harmful act probability distribution that used to be to the left of x^T (meaning exoneration) will now be to the right of x^T (meaning that there is liability). The greater that area—which is to say, the higher the curve—the larger the increase in $P_{HARMFUL}(x^T)$. (These alternative methods of assessing the change in the likelihood of liability are equivalent precisely because of the manner in which Figure 2 was derived from Figure 1, as explained previously.[45])

Given how much the probability of liability in adjudication rises, we can compute the increase in the expected sanction simply by multiplying by the other two factors: the likelihood of identification in stage one and the level of sanctions. Suppose that the expected sanction for those contemplating harmful acts rises from 100 to 110. That is, suppose that the product of the three factors was 100 initially and that $P_{HARMFUL}(x^T)$ rises by 10%.

We now need to determine how many individuals this increase in the expected sanction will deter. To answer this question, consider first which sorts of harmful acts will be deterred: they will be those that generate private benefits between 100 and 110,[46] for those are the acts whose private benefits exceeded the expected sanction initially (when it was at 100) but are now less than the expected sanction (110) when the evidence threshold is reduced. Acts with private benefits below 100 were already deterred, and those with benefits above 110 remain privately beneficial and hence continue to be undeterred.

The final step in computing how many individuals will be deterred by the reduction in the evidence threshold is to determine the concentration of marginal harmful acts, which is to say, how often private benefits from harmful acts fall in this range. The answer will depend on the context and on the level of the expected sanction at the outset. For example, it may be that most harmful acts were already deterred, private benefits usually being quite low, far below 100. In that case, very few individuals may have private benefits between 100 and 110. Or it may be that few are deterred, but most have private benefits well in excess of 100, and thus also in excess of 110, in which case

45. Formally, the height of each curve in Figure 1 is the absolute value of the derivative of the corresponding curve in Figure 2. *See supra* note 36.

46. Individuals whose acts would generate private benefits of precisely 100 were initially indifferent; after the reduction in the evidence threshold, those with private benefits of precisely 110 will be indifferent. Imagining the distribution to be continuous, it is inconsequential how such individuals will behave. *Cf. supra* note 34 (describing how, with a continuum, any particular real number has zero probability). One might suppose, for example, that indifferent individuals refrain from acting, or that they flip a coin.

nearly everyone will remain undeterred. If instead the initial expected sanction is close to the level of many individuals' private benefits—closer to the peak, or sweet spot, of the distribution of private benefits[47]—then the number deterred will be substantial.

For example, if many firms find the decision whether to (improperly) dump their waste into the nearby river to be a close call in light of the expected sanction, then reducing the evidence threshold would deter quite a few. By contrast, if most are already compliant or if nearly everyone flouts the law because the expected sanction is negligible relative to the private benefits, then a modest reduction in the evidence threshold would have little deterrent effect. Note, however, that if most are compliant at an expected sanction of 100, it may be that if the expected sanction were initially only 10, many would be near the margin, so reducing the evidence threshold in that case would matter a great deal. Likewise, if most flout the law at an expected sanction of 100, it might be that if the expected sanction were initially 500, perhaps much closer to the cost of proper disposal, then reducing the evidence threshold would deter many.[48]

Having determined how many individuals will be deterred and what this magnitude depends on, it remains to assess the net social gain per deterred act. As Figure 4 indicates, this social gain is the difference between the social harm of the act and the private benefit it generates. 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.

From the social harm per act, we need to subtract each act's private benefit. The magnitude of the private benefit of each forgone act equals the expected sanction that just suffices to deter it. Taking the above example in which reducing the evidence threshold raises the expected sanction on harmful acts

47. To be clear, the distribution of private benefits from acts is an entirely distinct concept from the distribution of evidence generated by acts; there need not and in general would not be any relationship between the two.

48. Some acts, perhaps price fixing or certain forms of tax evasion (each undertaken to a particular degree), may generate very similar levels of benefits to different actors. This relationship will tend to hold when the private benefit consists of profits and an act (of a given magnitude) is commonly associated with a particular profit level. In such cases, there may be little deterrence when the expected sanction is even moderately less than profits and nearly complete deterrence when the expected sanction is slightly higher than profits. Reducing the evidence threshold would thus be potent when the expected sanction is close to the level of profits but not otherwise. More realistically, there will often be greater heterogeneity among actors due to different reputational or other costs of violations, varying perceptions of how the legal system works, and differential actual likelihoods of being identified at stage one or found liable at stage two. The judgment-proof problem may also render a given sanction's impact nonuniform.

from 100 to 110, acts that are deterred each have private benefits of approximately (let us suppose) 105.[49] So, if the social harm of the act in question was 300, the net social gain per deterred act would be almost 200. Put concretely, if the private benefit of the harmful act is avoiding the higher production costs involved in compliance, then the net social gain from deterrence is not the full social harm avoided but only the difference between that level of harm and the cost that must be incurred to avoid it.[50] Note further that the net social gain per deterred act depends very much on how high the expected sanction is initially. For example, if the expected sanction begins at 275 rather than 100, the forgone private benefit per deterred act when this sanction rises from, say, 275 to 285, is approximately 280, and the net social gain from deterring an act is only about 20 rather than nearly 200. In other words, the magnitude of the social gain from deterrence depends on whether underdeterrence is initially substantial or modest.[51]

In summary, the deterrence benefit from a reduction in the evidence threshold depends on a number of factors. They are the increase in expected sanctions for harmful acts (the change in which depends on the degree to which the probability of liability for a harmful act, conditional on stage-one scrutiny, rises), the concentration of marginal harmful acts, and the magnitude of underdeterrence (which is the difference between social harm and the

49. For very small changes in the evidence threshold (for infinitesimal changes, when taking the derivative), the change in the expected sanction is small (approaching zero), and the range is extremely narrow (a point). For discrete changes, one would take the average of private benefits for acts that are deterred.

50. The assumption that the private benefits of individuals who commit harmful acts count as social benefits, providing some offset to the harm, may be controversial in some settings (notably, in the case of certain crimes; *see, e.g.*, KAPLOW & SHAVELL, *supra* note 2, at 320 n.54, 418-31). Note that the analysis could be modified by excluding this component and proceeding otherwise as indicated. However, the view that private benefits are indeed benefits to individuals and hence should be included does have appeal. And, in many specific contexts, inclusion is not regarded as controversial. For example, costs of performing contracts and employing more expensive technology to reduce pollution are real social costs that are generally counted, and individuals' benefits from many harmful acts—breaching contracts, polluting—are the avoidance of, or reduction in, such costs.

51. As mentioned in Subsection I.A.1, analysis is largely confined to the case in which the social harm exceeds the expected sanction and thus the private benefit of marginal acts: that is, where there is underdeterrence relative to the first best. Clearly, if the harm is less than the private benefit of the act, then each act deterred involves a net social loss. (For clarity of terminology, however, it is helpful to refer to such cases as involving deterrence—here, overdeterrence—reserving the term chilling for the discouragement of benign acts, which here are acts that involve no external social harm.) As will be apparent when considering chilling costs in a moment, in this case the evidence threshold should definitely be raised because doing so would be desirable on both fronts.

private benefit of a marginal harmful act, the latter of which equals the expected sanction).

Against this deterrence benefit, we must weigh the chilling cost. The right side of Figure 4 indicates that the chilling cost will equal the increase in the expected sanction for benign acts weighted by the concentration of marginal benign acts, which product is then multiplied by the net cost per act that is chilled – itself equal to the private benefit of chilled acts. Each of these elements needs elaboration, but the discussion will be brief because these components are now familiar.

In this case, the product of the first two terms expresses how many benign acts will be chilled when the evidence threshold is reduced. The increase in the expected sanction for benign acts is determined, qualitatively, in the same manner as for harmful acts, the difference being that now we are concerned with the magnitude of the increase in $P_{BENIGN}(x^T)$. As one slightly reduces x^T, we move a bit to the left in Figures 1 and 2. From Figure 2, the increase in the probability of a (mistaken) finding of liability is given by the (absolute value of the) slope of the $P_{BENIGN}(x^T)$ curve; or, from Figure 1, this magnitude is indicated by the height of the probability distribution curve for the benign act.

Using this rise in the probability of (mistaken) liability, we can determine the increase in the expected sanction for benign acts. Suppose that it rises from 10 to 12. The number who are newly chilled as a consequence will depend on how many individuals who contemplate benign acts have private benefits between 10 and 12. Acts with private benefits less than 10 were already chilled, and those with benefits above 12 remain privately advantageous and will continue to be unchilled. If there is a high concentration of private benefits in this intermediate region, many will be chilled, but if most have much lower benefits or far greater benefits, few additional acts will be chilled.

The net cost per chilled act is straightforward: it is simply the act's private benefit because the social loss consists of the forgone benefit from the act. These private benefits are also social benefits, and there is no external harm that is avoided when individuals desist from such acts. In the preceding example, the average benefit is, we may suppose, approximately 11. As with deterrence, this final component depends directly on the initial level of the expected sanction. If it was initially only 2 and would rise to 4, the net cost per chilled act would only be (roughly) 3, whereas if the initial expected sanction for benign acts was originally 60 and rose to 62, the cost per chilled act would be about 61. Note that, in this latter pair of cases, the ratio of the chilling costs per act is about 20 to 1 even though the magnitude of the increase in the expected sanction is the same. In other words, the chilling cost depends very strongly on the magnitude of the expected sanction for benign acts.

Multiplying these factors gives the overall chilling cost from a reduction in the evidence threshold. With one exception, the external social harm from harmful acts, the components are qualitatively the same as when computing the deterrence benefit, although the magnitudes of each will differ, often greatly, and in different directions. For example, the magnitudes of the increases in expected sanctions depend on the increases in $P_{HARMFUL}(x^T)$ and $P_{BENIGN}(x^T)$, which we see from Figures 1 and 2 can differ substantially, and in either direction.[52] (The magnitude of the increase also depends on the identification rates for the two types of acts in stage one—which may differ, often one supposes in the direction of a higher rate for harmful acts—and on the sanction, which is the same for the two acts.) The concentration of marginal acts also may differ significantly and in either direction, for it depends on how concentrated are the private benefits at the initial level of the expected sanction for the two types of acts, and these concentrations need not have any relationship to each other. (For example, benign acts could typically be highly privately beneficial compared to similar-looking harmful acts, or they might be much less so. Moreover, we measure this concentration at the initial levels of expected sanctions, which are generally greater for the harmful acts.) In addition, the net gain from deterrence subtracts the value of the private benefit, whereas the chilling cost is given by the positive value of the private benefit. Furthermore, the magnitude of each depends on the expected sanctions for each, which differ. Finally, for the deterrence benefit, we also include the social harm avoided, which has no analogue with regard to the chilling cost.

In all, we have focused on seven factors that determine whether a reduction in the evidence threshold from some initial level is desirable. (And, if it is undesirable, an increase in the threshold would be beneficial.) The optimum will have the feature that the deterrence benefit just equals the chilling cost—that is, there would be an equals sign between the two boxes in Figure 4.[53]

In thinking about this tradeoff, it is useful to note that, as one reduces the evidence threshold more and more, the magnitudes of the lower elements in

52. They are equal only at the point where the two probability distribution curves in Figure 1 cross. To the right of that point, $P_{HARMFUL}(x^T)$ rises more as x^T is reduced, and to the left, $P_{BENIGN}(x^T)$ rises more.

53. The text abstracts from some formalities. Notably, the posited equality is a necessary condition for what is referred to as an interior optimum, that is, an optimal x^T that is not extreme. Such an extreme would arise, for example, if social harm was negative (the external harm was actually a benefit), in which case no sanctions would be optimal and, given positive sanctions, an infinite evidence threshold would therefore be desirable. For a technical treatment, see Kaplow, Optimal Burden of Proof, *supra* note 16.

Figure 4 are falling for the deterrence benefit and rising for the chilling cost. As deterrence rises, the private benefit to individuals at the margin of committing harmful acts increases, which diminishes the net gain from deterrence; as explained above, this advantage is larger when underdeterrence is greater. Likewise, as chilling rises, the private benefit to individuals at the margin of committing benign acts similarly increases, which magnifies the chilling cost. Note that, when the risk of mistaken imposition of sanctions on benign acts is very low and thus the expected sanction on benign acts is quite small, the net cost per chilled act—equal to the private benefit, which, as explained, equals the expected sanction—will be negligible, favoring a lower evidence threshold as long as there is a nontrivial deterrence benefit. As a consequence of these factors, if, initially—say, at a very high evidence threshold—deterrence benefits exceed chilling costs, this inequality will tend to diminish and ultimately reverse as the evidence threshold is reduced. This point is not the entire story, however, because one must also attend to other factors; notably, the middle elements in each of the boxes may be rising or falling, which complicates the analysis.[54]

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.[55]

54. For example, although the net gain per deterred act is falling as the evidence threshold is reduced, it is possible that the concentration of marginal harmful acts is rising, and at an even greater rate, in which case the deterrence benefit in such a region would be rising. Furthermore, if we are in a range in which the deterrence benefit is rising, and at a faster rate than is the chilling cost, then one should continue reductions in the evidence threshold at least until this is no longer the case.

55. Note that greater harm favors a lower evidence threshold, but higher stakes more broadly have no immediate impact on the optimal threshold. To elaborate the latter point, suppose that, in moving from one context to another, external social harm and individuals' private benefits from both types of acts (more precisely, the distributions of private benefits for each) rise by a factor of ten—and that the sanction is also ten times as high. In such a case, the marginal harmful act and the marginal beneficial act initially have private benefits that are ten times as great. Then, only the bottom elements in Figure 4 will change, each rising by a factor of ten. (For the deterrence benefit, if both the social harm and the private benefit are multiplied by ten, the difference likewise rises tenfold.) Multiplying both sides by the same positive factor does not change the direction of inequality or the level of the evidence

The formulation represented in Figure 4 is, unfortunately, complex. Much more worrisome, it depends on a number of empirical facts that are likely to be quite difficult to ascertain and that, as just suggested, vary by context.[56] Nevertheless, it should be apparent from the foregoing exposition that each element really is central to the setting of an optimal evidence threshold. That is, if one wishes to design the legal system in a manner that advances social welfare, it is hard to see how one could avoid attending to how much changes in the evidence threshold affect the expected sanctions for harmful and benign acts, how many acts of each type are thereby deterred or chilled, and what is the social impact as a consequence of discouraging each of the two types of acts.[57] In addition, this exposition does, on reflection, appear to be reasonably complete in its domain[58] in that it includes the two fundamental behavioral

threshold at which the two sides would be in balance. The simple conclusion that greater harm, ceteris paribus, favors a lower threshold and the point that a proportional scaling of stakes and sanctions has no influence on the optimal threshold are contrary, for example, to Lord Denning's statement: "In criminal cases the charge must be proved beyond reasonable doubt, but there may be degrees of proof within that standard. Many great judges have said that, in proportion as the crime is enormous, so ought the proof to be clear." Bater v. Bater, [1950] 2 All E.R. 458 at 459 (Can. Que.); *see also* RONALD L. CARLSON ET AL., EVIDENCE: TEACHING MATERIALS FOR AN AGE OF SCIENCE AND STATUTES 922 (6th ed. 2007) ("In principle, the court's choice of a measure for the burden should reflect the stakes in the case; the more important the stakes, the higher the burden should be."). These points are also contrary to a common justification for the requirement of proof beyond a reasonable doubt in criminal cases. For example, in *In re Winship*, 397 U.S. 358, 363-64 (1970), the Supreme Court's rationale relies heavily on the high stakes for possibly innocent criminal defendants, but the court does not address why society's stakes regarding mistaken exoneration are not similarly high. Likewise, the Court also argues that the public's confidence in the legal system may be eroded by false convictions, *see id.* at 364, but does not mention that its confidence may also be undermined by frequent mistaken acquittals.

56. Note further that the elements have different natures. For the first element in each box, the variable that changes is the probability of liability, given that an individual is before the tribunal. For the two types of acts, this depends on the evidence that is produced—in particular, on how the relative degree of discrimination changes as the evidence threshold changes, as discussed in Subsection I.A.2. The second elements depend on something altogether different: how concentrated are individuals' benefits from the two types of acts, evaluated at the prevailing level of the expected sanction for each. For the third elements, the forgone private benefits are, as explained, equal to the expected sanction, which depends (also) on the stage-one identification rates for the two types of acts and on the sanction applied. And, for the deterrence benefit, the third element also depends on the external social harm, also qualitatively distinct from each of the other factors.

57. For some discussion of how the legal system might already attend to these considerations or might be adjusted to do so to a greater extent, see Section II.D.

58. As will be elaborated in Parts III and IV, it is not complete in all dimensions, such as with regard to administrative costs and the possible social costs of sanctions, which factors have

effects of changes in evidence thresholds and, with respect to each, considers the pertinent components. In sum, although the test for the optimal evidence threshold may initially seem formidable and foreign, it is on reflection quite intuitive, combining plausible elements in a natural manner.[59]

II. CONVENTIONAL CONCEPTIONS COMPARED

A. Conventional Burden of Proof

In elucidating conventional burden of proof notions and especially when comparing the analysis here to that in Part I, on optimal evidence thresholds, it is useful to adhere to the same framework. Accordingly, as in Section I.A, we will continue to examine a setting in which some individuals may commit harmful acts, others may commit benign acts, and a stage-one process results in a given fraction of each type of committed act coming before a tribunal. In stage two, adjudication, the factfinder's problem in any given case is, as before, to decide whether to hold the person liable and apply the designated sanction or, instead, to exonerate the individual.[60]

Under conventional burdens of proof, the factfinder is supposed to assess liability if the likelihood that the individual committed the harmful act is sufficiently high. (To be clear, the present treatment considers conventional proof burdens in their idealized form, assuming for the sake of analysis that they are taken seriously—specifically that they are followed to the letter. Section II.D revisits the subject, suggesting that actual practice may deviate in important respects.) For concreteness, the analysis usually considers the preponderance of the evidence standard that is ordinarily employed in civil cases in the United States, under which liability is found if and only if it is more likely than not that the case involves an individual who committed the harmful type of act rather than the benign type.[61] At the end of this Section, it

intentionally been set aside in this first-pass analysis that is designed to concentrate on the effects of the legal system in influencing individuals' behavior.

59. In addition, as Subsection II.B.2 elaborates, the 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.

60. Additionally, as mentioned in note 24, settlement and plea bargaining are ignored. Regarding the present analysis, however, which concerns the preponderance rule and others based on ex post likelihoods, accounting for settlements and pleas would further complicate the factfinders' decisionmaking, as explained in note 218.

61. *See supra* note 3 (on the probabilistic interpretation of the preponderance rule).

will be explained how the analysis is qualitatively the same for any requisite likelihood.[62]

To apply this burden of proof concept, it is necessary to determine the likelihood that an individual before the tribunal committed the harmful type of act. (Note that this likelihood did not need to be assessed in Part I for purposes of determining the optimal evidence threshold, a contrast discussed further in Sections II.B and II.C.) In any given case, there will be some set of evidence, the strength of which we continue to denote by x. The pertinent likelihood in such a case will depend on two factors: the relative frequency that the acts coming before the tribunal are of the harmful type, and the relative probabilities that the two types of acts will generate evidence of strength x.

Begin by supposing (unrealistically in most instances) that harmful and benign types of acts come before the tribunal with equal frequency. In that event, in a given case with evidence x, it will be more likely than not that the act is of the harmful type when the harmful act is more likely to generate evidence x than is the benign type of act. To take a simplified example, suppose that there are 200 cases, 100 with harmful acts and 100 with benign acts. Furthermore, harmful acts generate evidence of x 10% of the time and benign acts generate such evidence only 5% of the time.[63] Then we would expect 10

62. There is some controversy surrounding probabilistic interpretations of the burden of proof, particularly in the criminal context where, in the United States, it is required that proof be beyond a reasonable doubt. *See infra* note 78. This Article is not directed to the interpretive controversy, but it is worth noting that the concept of subjective probability and related demands of logical consistency indicate that decisionmakers should behave as if following some probabilistic rule. *See* LEONARD J. SAVAGE, THE FOUNDATIONS OF STATISTICS 6-104 (1954). To see this, consider a set of cases that are identical except for the probability that the individual before a tribunal committed a harmful act. Take a case in which that probability is p_1 and suppose that, after reflection, the decisionmaker decides that it is appropriate for liability to be imposed. Then consistency demands the imposition of liability in all cases in which the probability exceeds p_1 (for, after all, it was stipulated that the cases in the set differ only in their probabilities). In another case, the probability is at some lower level p_2, and the decisionmaker decides that it is appropriate to exonerate. Then consistency demands exoneration in all cases in which the probability is less than p_2. As one considers additional cases, with probabilities between p_1 and p_2, further decisions will narrow the undecided range. Ultimately, there will, in principle, exist some probability p^* above which the decisionmaker would always find liability and below which it would always exonerate. (If not, then there must exist at least one pair of cases with differing probabilities—and all else equal—such that the decisionmaker would exonerate in the case with the higher probability but find liability in the case with the lower probability, which does not make sense.) For purposes of the present analysis, as just stated in the text, nothing turns on what the target probability p^* is; all that is required is consistency.

63. For ease of exposition, the examples in this Section use discrete probability distributions even though the preceding figures use continuous ones.

cases (10% of 100) involving harmful acts to generate *x* and 5 cases (5% of 100) with benign acts to generate *x*. In sum, 10 in 15 cases (two-thirds, approximately 67%) that generate *x* will be ones with harmful acts. Therefore, in a case with evidence of strength *x*, it is more likely than not that it is one involving a harmful type of act. By contrast, if harmful acts generate evidence of some other, much lower level of *x* only 2% of the time but benign acts generate that low level 8% of the time, then in only 2 in 10 (20%) of the cases will a tribunal that observes this lower level of *x* have before it an act of the harmful type. Such a case fails to satisfy the requirement that it be more likely than not that the act is of the harmful type.

The probability of evidence at any level *x*, given the type of act, can be determined from Figure 1 in Subsection I.A.2. Specifically, this probability is given by the height of the pertinent curve.[64] For high levels of *x* (to the right of center in the figure), harmful acts are more likely to generate the evidence, and for low levels of *x* (to the left of center in the figure), benign acts are more likely to generate the evidence. The likelihoods are equal where the two curves cross. And, more broadly, the relative likelihood that the harmful act will generate evidence of strength *x* is given by the ratio of the height of the probability distribution curve for harmful acts to the height of the curve for benign acts, each measured at the *x* in question.[65]

The analysis to this point is incomplete, for it begins with the simplifying assumption that the two types of acts come before the tribunal with equal frequency in the first place.[66] In general, this will not be so. For example, if most harmful acts are deterred and there are a large number of benign acts that may be confused with harmful ones, then the bulk of cases coming before the tribunal will be ones involving benign acts. (Some argue this to be the case under competition law with regard to predatory pricing allegations, where most suits are said to be brought by less efficient competitors complaining

64. It is common to refer to the height of the curves in Figure 1 as indicating probabilities although, as discussed in note 34, the probability of evidence being exactly at some level *x* (with either type of act) is zero with a continuous distribution. The height is also sometimes called the frequency or the density. What matters for the analysis, as emerges in the text, is the ratio of these heights, which remains a meaningful concept.

65. Indeed, the very fact that this likelihood ratio is higher for larger values of *x* is why it makes sense to refer to *x* as a (positive) indicator of the strength of the evidence.

66. The point that base rates are also relevant in Bayesian analysis (see the discussion in the text later in this Subsection) and can have dramatic effects is routinely emphasized in elementary discussions. *See, e.g.*, *Bayes' Theorem*, WIKIPEDIA, http://en.wikipedia.org/wiki/Bayes%27_theorem (last visited Sept. 5, 2011); *False Positive Paradox*, WIKIPEDIA, http://en.wikipedia.org/wiki/False_positive_paradox (last visited Sept. 5, 2011).

about tough but actually legitimate competitive behavior.[67]) By contrast, if deterrence is weak and there are very few benign acts that might readily be confused with harmful acts (or if most that might be so confused are chilled in any event), then most acts before the tribunal will be harmful ones. (Perhaps this was the case with drunk driving when penalties were low and enforcement was lax.[68]) These differences directly influence the frequency assessment.[69]

To illustrate this point,[70] suppose that instead of 100 harmful acts and 100

67. *See, e.g.*, Andrew I. Gavil, *Burden of Proof in U.S. Antitrust Law*, *in* 1 AM. BAR ASS'N SECTION OF ANTITRUST LAW, ISSUES IN COMPETITION LAW AND POLICY 125, 134-36 (Wayne Dale Collins et al. eds., 2008) (describing and criticizing this view). U.S. courts have imposed strong burdens on predatory pricing claims that they justify in large part on the ground that most such cases flowing into the legal system are likely to involve benign rather than harmful behavior. *See, e.g.*, Brooke Grp. Ltd. v. Brown & Williamson Tobacco Corp., 509 U.S. 209, 227 (1993); Matsushita Elec. Indus. Co. v. Zenith Radio Corp., 475 U.S. 574, 588-90 (1986). For further discussion, see Louis Kaplow & Carl Shapiro, *Antitrust*, *in* 2 HANDBOOK OF LAW AND ECONOMICS 1073, 1196-97, 1199-1200 & n.192 (A. Mitchell Polinsky & Steven Shavell eds., 2007).

68. Regarding the latter, if police tended to stop only vehicles that exhibited extremely erratic behavior and arrest only those who badly failed field sobriety tests, then most coming before a tribunal would be individuals who had engaged in drunk driving. Confused benign acts might be individuals who were temporarily dizzy for idiosyncratic reasons, were observed to swerve wildly, and, when pulled over, could not maintain their balance. A factfinder's assessment of liability (in a criminal setting, by a higher standard of proof) would logically be based, in significant part, on how infrequent would be those acts that might be confused with extreme drunk driving in the first place.

69. From the analysis in Part I, we know that the frequency with which acts of each type come before the tribunal will be determined by individuals' opportunities to commit the two types of acts and the level of the expected sanction for each type. For the harmful act, the expected sanction is given by the product of the stage-one identification rate, the probability of liability (itself a function of the prevailing evidence threshold) in light of the distribution of evidence generated by harmful acts, and the sanction. All individuals with opportunities to commit harmful acts that yield benefits greater than this expected sanction will commit them, and the fraction of those committing harmful acts that come before the tribunal will depend (again) on the stage-one identification rate for harmful acts. The frequency for benign acts is determined analogously.

70. Alternatively, consider cases in which x is at the level where the two probability distribution curves in Figure 1 cross, that is, where each type of act is equally likely to generate evidence of x. Then, if the majority of the cases entering adjudication involve harmful acts, the likelihood that a given case with evidence of x involves a harmful act will exceed 50% and thus satisfy the preponderance rule; if the majority involve benign acts, then the likelihood will be less than 50% and thus fail to satisfy the preponderance rule. In particular, if almost no one committed the benign type of act (but many committed the harmful type of act), the preponderance rule would be satisfied, but if almost no one committed the harmful type of act (but many committed the benign one), the preponderance rule would not be satisfied. In sum, even though, *conditional on the type of act being before the tribunal*, the likelihood of generating this value of x is equal for the two types of acts, the likelihoods that a case generating x is of one versus the other type are not in general equal. Equality holds only if there is an equal number of each type of case flowing into the system.

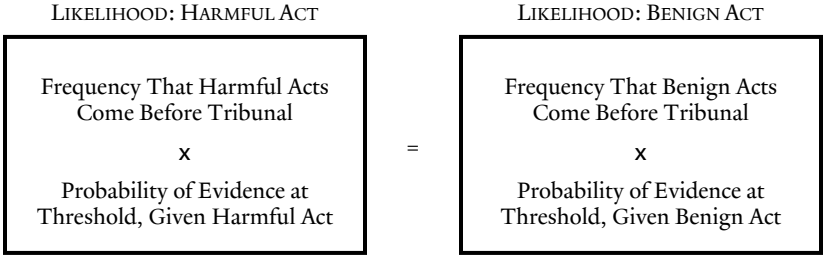
benign acts, we have 20 harmful acts and 100 benign acts. Then, if (as in our first example above) harmful acts generate evidence of x 10% of the time and benign acts generate x 5% of the time, we would expect that there will be 2 cases (10% of 20) in which evidence x was generated by a harmful act and, as before, 5 cases in which it was generated by a benign act. Thus (rather than a two-thirds likelihood that an act generating x is of the harmful type, as before), there would be a two-sevenths likelihood that a harmful act is before the tribunal, which would call for exoneration under the preponderance rule—even though harmful acts generate evidence of x with twice the likelihood that benign acts do. Now, suppose that we have 200 harmful acts and 25 benign ones and (as in our second example above) harmful acts generate evidence of x 2% of the time and benign acts generate it 8% of the time. Then a tribunal will expect to be confronted with 4 (2% of 200) cases in which the acts are harmful and 2 (8% of 25) in which the acts are benign, so the preponderance rule will be satisfied—even though benign acts generate evidence of x with four times the likelihood that harmful acts do.[71]

Having explained the two factors that bear on whether the preponderance rule is satisfied in a given case with evidence of strength x, we can now present an explicit formulation of the evidence threshold, x^T, that implements the preponderance rule. We are asking what level of x^T is such that it is equally likely that we are seeing a person who committed a harmful act versus one who committed a benign act. Or, put another way, we are determining the value of x^T such that it is more likely than not that a tribunal presented with evidence x is facing a harmful act if and only if $x > x^T$.[72]

The relationship that characterizes the evidence threshold x^T consistent with the preponderance rule is depicted in Figure 5.

Figure 5.

EVIDENCE THRESHOLD FOR THE PREPONDERANCE RULE



71. For elaboration on these results, which initially may seem counterintuitive to some, see Subsection II.C.1.

72. Subsection II.C.1 presents important complications.

The product in the box on the left side indicates how often harmful acts before the tribunal will be associated with evidence at the threshold x^T, and the product in the right box indicates how often this is so for benign acts. If the product in the left box exceeds that in the right box, then the acts are more likely to be harmful; in this case, it seems that the x^T that satisfies the preponderance rule must be lower. If the product on the left is less than that on the right, then the acts are more likely to be benign; in this case, it seems that the requisite x^T must be higher. The evidence threshold that satisfies the preponderance rule is that where these two likelihoods are equal; that is, when the evidence strength is precisely at such an x^T, it is just as likely that the case at hand involves a harmful act as a benign one.

The two elements in each box correspond to the two factors identified and illustrated previously. The frequencies that harmful and benign acts come before the tribunal are the rates at which each type of case flows into the system of adjudication. The probabilities of evidence at the threshold, given the type of act, are the heights of the probability distribution curves in Figure 1. The preceding numerical examples illustrate the manner in which these factors interact to determine each likelihood and thus whether the preponderance rule is satisfied.

Note that if the two frequencies (the top elements in each box) were equal, then the preponderance rule threshold would be one for which the probabilities of evidence of that strength were equal for the two types of acts (which is the level of x where the two probability distribution curves in Figure 1 cross, at the center of the axis). If the frequency of harmful acts is relatively greater than that for benign acts, then the evidence threshold under the preponderance rule will be such that the probability of evidence at the threshold when a benign act was committed is correspondingly higher than that for a harmful act (the value of x^T would be lower, to the left of center in Figure 1). The reason is that—for the products in the two boxes to be equal—if the first factor is higher in the left box, then the second factor in the left box must be lower (and by the same proportion). Similarly, if the frequency of harmful acts is relatively less, then the evidence threshold under the preponderance rule will be such that the probability of evidence at that level for

the harmful act is relatively greater (the value of x^T would be higher, to the right of center in Figure 1).[73]

The numerical examples presented before introducing Figure 5 depict this phenomenon. Moreover, these results are intuitive: If most acts before tribunals are harmful, then the value of x^T such that the likelihood that the act before the tribunal is a harmful one is only 50% will be quite low (strong evidence would merely reinforce the prospect that the act under adjudication was a harmful one, making that prospect much more likely than not). If most acts before tribunals are benign, then the value of x^T that is necessary for the likelihood to reach 50% would be rather high (it would take quite strong evidence to overcome the initially low prospect that the act is of the harmful type).

Some readers will recognize the explanation of Figure 5 as standard Bayesian inference.[74] The frequencies of harmful and benign acts coming before the tribunal determine the so-called prior probability that an act is of the harmful type,[75] that is, the probability before looking at the evidence that the case at hand involves a harmful act. The probabilities of evidence at the threshold given the type of act are used for updating purposes, that is, to transform the prior probability into the actual probability in light of the

73. Viewing Figure 5 as an equation, we can divide both sides by the probability of evidence at the threshold, given a benign act, and also divide both sides by the frequency that harmful acts come before the tribunal, which is depicted in the following figure:

RATIO OF PROBABILITIES		RATIO OF FREQUENCIES
Probability of Evidence at Threshold, Given Harmful Act ÷ Probability of Evidence at Threshold, Given Benign Act	=	Frequency That Benign Acts Come Before Tribunal ÷ Frequency That Harmful Acts Come Before Tribunal

On the right side, we now have the ratio of the frequency of benign acts coming before the tribunal to the frequency of harmful acts. On the left side we have the ratio of the probability of evidence at the threshold, given a harmful act, to this probability for a benign act. The equal sign informs us that these ratios must be equal for x^T to be consistent with the preponderance rule.

74. *See, e.g.*, HOWARD RAIFFA, DECISION ANALYSIS: INTRODUCTORY LECTURES ON CHOICES UNDER UNCERTAINTY 17-19, 30-31 (1968).

75. Specifically, this prior probability equals the frequency of a harmful act coming before the tribunal divided by the sum of that frequency and the frequency for the benign act.

evidence.[76] This resulting value is referred to as the posterior probability that the act before the tribunal is the harmful type. This result using Bayes' Theorem is essentially what we have just presented; in particular, the posterior probability equals the value of the product in the left box divided by the sum of the products in the left box and in the right box. The evidence threshold x^T for the preponderance rule is that value of x for which this fraction (posterior probability) is 0.5 or 50%.

Before proceeding, it is important to recognize that essentially the same analysis and a slightly modified version of Figure 5 can be used to determine the evidence threshold consistent with any desired minimum likelihood that the act before the tribunal is of the harmful type. To illustrate, suppose that one wanted to require a likelihood of at least 75% rather than 50% (this elevated likelihood might roughly be associated by some with the "clear and convincing" standard[77]). Then one could simply multiply the product in the right box by 3. To see why, note that our target likelihood that the act under scrutiny is a harmful one is 75%, which implies that the likelihood that it is benign is 25%. To achieve this relationship, we need the value in the left box to be 3 times higher than the value in the right box (75% = 3 x 25%). If we multiply the value on the right side by 3 and then determine for what evidence threshold x^T there is equality, we will have our answer. If we instead wanted a 95% likelihood that the act is of the harmful type (which some might loosely associate with the beyond a reasonable doubt requirement[78]), we would need

76. The prior probability is, according to Bayes' Theorem (or Rule), multiplied in this instance by the probability that evidence is of a given strength when the act is the harmful type (the conditional probability), divided by the overall probability that evidence of the given strength will arise (the marginal probability, which in turn equals the sum of the conditional probability for the harmful type and that for the benign type, each conditional probability being weighted by the corresponding prior probability).

77. *See, e.g.*, SYDNEY A. BECKMAN, SUSAN CRUMP & FRED GALVES, EVIDENCE: A CONTEMPORARY APPROACH 697 (2009); C.M.A. McCauliff, *Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?*, 35 VAND. L. REV. 1293, 1328 tbl.5, 1332 tbl.8 (1982) (presenting results from a survey of federal judges indicating that the standard was associated with a probability of 75% by just over a third, with either 70% or 80% by nearly a third more, and with 74.99% on average).

78. Some embrace probabilistic interpretations of all proof standards. *See, e.g.*, 2 MCCORMICK ON EVIDENCE, *supra* note 3, at 490-91 (stating, when commenting on the requirement of proof beyond a reasonable doubt, "As we have seen with reference to civil cases, a lawsuit is essentially a search for probabilities. A margin of error must be anticipated in any such search."). However, many commentators resist the idea of assigning any probability to the requirement of proof beyond a reasonable doubt, and there is considerable variation in what probabilities are thought to be associated with this proof standard when values are assigned. *See, e.g.*, RICHARD O. LEMPERT, SAMUEL R. GROSS & JAMES S. LIEBMAN, A MODERN APPROACH TO EVIDENCE: TEXT, PROBLEMS, TRANSCRIPTS AND CASES 1244 n.13 (3d ed. 2000)

the left side to be 19 times the value of the right side (95% = 19 x 5%), so we would multiply by 19 on the right. And so forth.[79]

Observe that multiplying the right side by a constant does change the value of the evidence threshold x^T but does not change the analysis by which it is determined. We still have the two factors in each box in Figure 5, and these two factors are evaluated as just described. Thus, although different strengths of the burden of proof are implemented by different evidence thresholds, the relevant elements and basic concepts are entirely the same—except for incorporating the multiplicative factor, which for the preponderance rule was implicitly taken to equal 1 (50% = 1 x 50%) and thus could be omitted from the figure. Accordingly, in most of the discussion to follow, the text will continue to speak, for concreteness, in terms of the preponderance rule. We can, however, interpret these references for most purposes as a shorthand for preponderance-like rules, which is to say rules of the class that the determinants of the evidence threshold are essentially the same as those under the preponderance rule—which includes all rules that deem it necessary and sufficient that the likelihood that the act is of the harmful type exceed some given likelihood (be it 50%, 75%, 95%, or any other value).

(discussing studies suggesting that most jurors place the requirement in the 85% to 90% range); MARC L. MILLER & RONALD F. WRIGHT, CRIMINAL PROCEDURES: CASES, STATUTES, AND EXECUTIVE MATERIALS 1319 (3d ed. 2007) ("Most state and federal courts discourage or prohibit trial judges from efforts to quantify the reasonable doubt standard."); *id.* (noting one study suggesting that jurors view beyond a reasonable doubt to mean about 86% certainty and another indicating a wide range from 50% to 100%); Hal R. Arkes & Barbara A. Mellers, *Do Juries Meet Our Expectations?*, 26 LAW & HUM. BEHAV. 625, 630-31 (2002) (citing surveys of judges, where mean and median numerical responses were at or near 90%, with 90%, 95%, and 100% being the most frequent specific responses); Erik Lillquist, *Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability*, 36 U.C. DAVIS L. REV. 85 (2002) (arguing that the reasonable doubt standard in action requires less certainty than we generally assume, and that jurors should and do use the standard flexibly); McCauliff, *supra* note 77, at 1325 tbl.2, 1332 tbl.8 (presenting results of a survey of federal judges showing that about a third associated beyond a reasonable doubt with 90%; most others offering 80%, 85%, 95%, or 100%; and a mean response of 90.28%); Charles R. Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 HARV. L. REV. 1187, 1196-97 (1979) (claiming that ambiguity in the meaning of the reasonable doubt requirement usefully serves the legal system's need to make verdicts externally acceptable to the public at large); Lawrence M. Solan, *Refocusing the Burden of Proof in Criminal Cases: Some Doubt About Reasonable Doubt*, 78 TEX. L. REV. 105 (1999); *see also infra* Section II.D (questioning the relevance of quantification of conventional proof requirements in a vacuum). Regarding the theoretical question whether the beyond a reasonable doubt requirement (or any other not-explicitly-quantitative standard) can be reduced to a probability, see note 62.

79. The formula that determines the factor by which to multiply on the right side equals the target percentage divided by the quantity: one minus the target percentage.

B. Comparison

This Section compares the determinants of the evidence threshold for the preponderance rule (and, as just noted, for any rule formulated in terms of the likelihood that the act before the tribunal is of the harmful type) with those of the optimal evidence threshold, which is to say that which maximizes social welfare. This being a conceptual analysis, the focus is qualitative, on how and why the relevant factors differ.

1. Analysis

To facilitate a side-by-side comparison, it is helpful to reproduce both Figure 5 and Figure 4, the latter (now denoted Figure 6) slightly altered by placing an equals sign between the boxes to denote the optimal threshold.

Figure 5.
EVIDENCE THRESHOLD FOR THE PREPONDERANCE RULE

LIKELIHOOD: HARMFUL ACT		LIKELIHOOD: BENIGN ACT
Frequency That Harmful Acts Come Before Tribunal x Probability of Evidence at Threshold, Given Harmful Act	=	Frequency That Benign Acts Come Before Tribunal x Probability of Evidence at Threshold, Given Benign Act

Figure 6.
EFFECTS OF A REDUCTION IN THE EVIDENCE THRESHOLD AT THE OPTIMAL THRESHOLD

DETERRENCE BENEFIT		CHILLING COST
Increase in Expected Sanction for Harmful Acts x Concentration of Marginal Harmful Acts x Net Gain per Deterred Act (Social Harm - Private Benefit)	=	Increase in Expected Sanction for Benign Acts x Concentration of Marginal Benign Acts x Net Gain per Chilled Act (Private Benefit)

On their face, Figures 5 and 6 have no elements in common. Figure 6, for the optimal evidence threshold, has seven elements: three of them—the increase in the expected sanction, concentration of acts, and private benefit—applicable to each type of act; and social harm, relevant for only the harmful act. Figure 5 has four elements: two—the frequency of acts and probability of evidence at the threshold—applicable to each type of act. The overlap is nil. (If we modified the right side of Figure 5 to include a constant multiplier to reflect a likelihood-style rule not pegged at 50%—for example, a factor of 3 for a 75% rule—we would be introducing yet another difference.)

On further analysis, it turns out that there are some underlying commonalities, although not nearly enough to alter the conclusion that the differences are overwhelming. Specifically, the top elements in Figure 6 for the optimal evidence threshold—the increases in expected sanctions—are determined in part (recall from Section I.B) by the extent to which the probability of finding liability, given that an individual is before the tribunal, rises as the evidence threshold is reduced. That change in probability for the two types of acts is, in turn, indicated by the height of the corresponding probability curves in Figure 1. (To review, as the evidence threshold x^T is reduced slightly, some of the area under the probability curves that used to be to the left of x^T will now be to the right of x^T, and the heights of the probability curves indicate how great these changes will be.) In Figure 5, for the evidence threshold under the preponderance rule, the bottom elements in the boxes are the probabilities of evidence at the threshold for the two types of acts. Those probabilities are also given by the heights of the curves in Figure 1. Accordingly, one component of the top elements in the boxes in Figure 6 is equal to the corresponding values of the bottom elements in the boxes in Figure 5.[80] No other direct linkages exist, although there are some further indirect connections, as noted in the margin.[81]

80. Interestingly, although these values are the same, they arise for different reasons. In Figure 6, for the optimal threshold, the origin pertains to the changes in deterrence and chilling as the threshold is reduced, whereas in Figure 5, for the preponderance rule, it pertains to the likelihood that evidence would be observed.

81. This is hardly surprising because there is a single underlying scenario used to determine all of the values in both figures. Specifically, the upper elements in the boxes in Figure 5, the frequencies with which each type of act comes before the tribunal, depend on the level of the expected sanction for each type of act, which level (as previously explained) also equals the private benefit of the marginal deterred or chilled individual. That private benefit is relevant in Figure 6 but not in any direct way in Figure 5. Additionally, one of the components in the first element in the boxes in Figure 6 (the increase in expected sanctions) is the fraction of each type of act that is identified in stage one, and this fraction is also a component of the first element in the boxes in Figure 5 (the frequency that acts come before the tribunal). For related discussion, see Subsection II.B.2.

To appreciate how great the difference is between these two approaches to setting an evidence threshold—choosing a target likelihood versus maximizing social welfare—it is useful to examine briefly each of the many differences between these two figures. The top elements in the boxes in Figure 6 are the increases in expected sanctions for harmful and for benign acts. If the concern is welfare, these factors are crucial because they indicate how much deterrence and chilling rise when the evidence threshold is reduced. They are not directly reflected in Figure 5, for the preponderance rule, because that rule is not concerned at all with the effects of the evidence threshold on behavior; concomitantly, it is concerned with properties *at a given level of the threshold* and not at all with how anything changes *as the threshold is changed*. This difference in perspective is fundamental and explains some of the other contrasts as well.

The middle elements in the boxes in Figure 6 are the concentrations of marginal harmful and benign acts. As explained in Section I.B, this information indicates whether the increases in expected sanctions for both harmful and benign acts translate into large or small numbers of individuals being deterred and chilled, respectively. Again, these factors are obviously central in assessing the welfare consequences of different evidence thresholds, but they have nothing to do with the likelihood, at some given evidence strength x, that the act before the tribunal is of the harmful rather than the benign type.

For the final row of Figure 6, we have individuals' private benefits of acts that are forgone when such acts are deterred or chilled, and the social harm that is avoided when harmful acts are deterred. These are the social welfare impacts of the changes in behavior just described. They are of obvious social consequence but again are entirely irrelevant to whether evidence of some strength x indicates that the likelihood of an act before the tribunal is above or below 50% (or any other stipulated magnitude).

Relatedly, the elements in Figure 5, for the preponderance rule, do not appear directly in Figure 6, and the reason is that they do not directly relate to the welfare effects of different evidence thresholds. The top elements in the boxes in Figure 5 are the frequencies with which acts of the two types come before the tribunal. Although of central relevance in computing the likelihoods, as explained in Section II.A, they have no direct welfare relevance in setting the optimal evidence threshold. This may be surprising because the frequencies of harmful and benign acts seem highly pertinent. However, as mentioned just above, what matters in deciding whether the evidence threshold is too high, too low, or just right (optimal) is how much the number of harmful and benign acts *changes* as the threshold is *changed*, not on how many acts of each type there are. (As will be discussed in Parts III and IV, however, in other settings and for other reasons, these overall frequencies of harmful and benign

acts will matter in determining the optimal evidence threshold.) Last, as already discussed, the bottom elements in the boxes in Figure 5—the probabilities of evidence at the threshold for the two types of acts—are the other key determinants of the likelihood, which is the matter of interest under the preponderance rule (and other likelihood-based rules) but is not relevant per se for determining the optimal evidence threshold (although it is a component of the change in expected sanctions for the two types of acts).

The large—one might say radical—difference between the two formulations has a number of implications. First, as a conceptual matter, it reinforces the notion that we do not just have some subtle or technical distinctions but rather two entirely different outlooks on what the burden of proof is all about. Figure 6, for the optimal evidence threshold, was derived by asking what would be the effects of changing the threshold on social welfare, and then analyzing how the relevant effects, the deterrence benefit and the chilling cost, are determined. By contrast, Figure 5, for the preponderance rule, was derived by asking under what circumstances would the likelihood that the case before the tribunal involves a harmful act be 50% (or some other stipulated probability).

Second, as a pragmatic matter, if one is actually concerned with welfare, taking a likelihood-based approach such as that under the preponderance rule does not serve as a proxy for identifying a welfare-maximizing evidence threshold, not even approximately so. A good proxy ordinarily is determined by most of the same elements as the true objective and in a similar fashion—or, relatedly, there needs to be a substantial correlation between the proxy and the objective. The foregoing comparison makes it plain that, instead, the two calculuses are almost entirely distinct. For example, it would be a pure coincidence if the evidence threshold for the preponderance rule was equal to or even close to the optimal level. After all, factors that could make the optimal evidence threshold very high or quite low have no impact at all on the evidence threshold that implements the preponderance rule, and vice versa.

An immediate corollary of this practical point is that the optimal evidence threshold could be associated with any ex post probability whatsoever. This claim is true in light of three considerations: First, the optimal evidence threshold is determined by the combination of a large number of factors, as depicted in Figure 6, that each could take on a wide range of values. Second, the ex post probability (which, recall from the discussion of the Bayesian posterior probability at the end of Section II.A, is the value of the product in the left box in Figure 5 divided by the sum of the products in the two boxes in Figure 5) is also determined by many factors that each could take on a wide range of values. Third, the determinants of the first and of the second are

largely unrelated, making it entirely plausible that the first could be high and the second low, the first low and the second high, and so forth.

To be more specific, consider the first point, regarding the optimal evidence threshold. All else equal, if harm is extremely high and few beneficial acts would be chilled, the optimal evidence threshold would be very low. However, if harm is negligible and many beneficial acts would be chilled, the optimal evidence threshold would be very high—one could imagine cases in which the optimum would be little different from de facto legal immunity. These perhaps jarring illustrations should not, however, be seen as particularly counterintuitive. Even in the analytically simpler medical diagnosis example mentioned in the Introduction and in Subsection I.A.2, it is familiar that the optimal treatment threshold may be extremely low or quite high depending on the risks and benefits of the treatment.[82] For example, an immunization may be advantageous to an individual with less than a one in a thousand chance of contracting a disease in light of the magnitude of harm averted and the small cost; by contrast, a dangerous experimental treatment may be undesirable even for an individual with a ninety percent chance of dying without it if the prospective benefit is sufficiently small and there is a serious risk that it would cause death in the remaining ten percent of cases. On reflection, we should instead be troubled by the notion that it may make sense, even as an approximation, to employ a single threshold or to demand a particular likelihood (such as fifty percent) to make important decisions in a wide range of contexts in which the consequences vary dramatically.

Now consider the second point concerning the evidence threshold implied, say, by the preponderance rule. One might suppose that it is at some intermediate, moderate level (perhaps near where the probability distribution curves in Figure 1 cross). However, as explained in Section II.A, this need hardly be the case due to the importance of the upper elements in the boxes in Figure 5 pertaining to the frequencies with which harmful and benign acts come before the tribunal. Suppose, for example, that many of the harmful type of act are committed but there are few benign acts that might be confused with it. In that case, a moderate evidence threshold would be associated with a very high likelihood that a given act before the tribunal was of the harmful type. To implement the preponderance rule may then require an extremely low evidence

82. This setting is closer to that examined in Section IV.C, where it will be explained that, even though the qualitative difference between the formulation for the optimal evidence threshold and that for the preponderance rule is far less, it is still true that the optimal threshold could take on any value and be associated with any ex post likelihood.

threshold.[83] Now imagine instead that there are many benign acts and few harmful ones. Then, in order to raise the ex post likelihood from a very low level to 50%, it may be necessary to set the evidence threshold at an extremely high level.[84]

In sum, optimal evidence thresholds may be very high or quite low. Likewise for the evidence thresholds required to implement the preponderance rule—or any other rule that targets an ex post likelihood that the individual before the tribunal committed the harmful type of act. And, as previously elaborated, there is little relationship between the factors determining the evidence threshold under these two contrasting conceptions of the problem, so indeed any relationship is possible. If one is concerned with social welfare, this finding should be disturbing if the legal system in fact operates under the preponderance rule or some other likelihood-based formulation,[85] for such a regime would sometimes insist on very high thresholds when the optimal threshold was not merely lower but very low, and it would sometimes employ quite low thresholds when the optimal threshold was extremely high.

2. Comment on Information Requirements

An obvious practical challenge posed by the formulation for the optimal evidence threshold developed in Part I and revisited in the preceding Subsection concerns the high information requirements for its application, whether to a broad class of cases, a narrower cluster, or a given case.[86] This concern is a serious one given the number of elements in Figures 4 and 6 as well as the difficulty of ascertaining many of them, such as the likelihoods that harmful and benign acts each give rise to evidence of a particular strength and the concentration of individuals' private benefits from acts of the two types.

83. The reader may wonder whether even this would suffice, for at a low threshold there may still be relatively few benign acts (there were not many to begin with, and the reduction in the threshold would chill even more). This concern turns out to be warranted, as discussed further in Subsection II.C.1.

84. Echoing the preceding footnote, it is not obvious that one would ever achieve the balance of likelihoods required by the preponderance rule.

85. As discussed in Section II.D, it is hardly clear that the legal system in fact operates in this manner.

86. The optimal level of generality of burden of proof rules is beyond the scope of the present investigation. It should be apparent, however, that both optimal evidence thresholds and the evidence threshold for the preponderance rule (or other rules of that type) will vary greatly by context, even at fairly refined levels. For example, even within contract disputes, different claims in different settings may vary substantially with regard to all of the key factors. For further discussion, see note 128.

This Subsection extends the comparison between an optimal evidence threshold and that implied by the preponderance rule (or others similarly grounded in ex post likelihoods) to consider differential information requirements.

Before examining the details, it should be emphasized at the outset that possible differences in information requirements have little to do with the merits of one or another way of setting the burden of proof. The optimal evidence threshold is that which maximizes social welfare. If welfare is the objective, it would make no sense to substitute some other formulation because it happened to be easier to employ when that alternative has essentially no relationship to what matters. (Easiest of all would be to flip a coin, but its ease offers little to commend it.) As emphasized in Subsection II.B.1, a preponderance rule does not depend at all on most of what is welfare relevant and instead depends on much that is not. That is, conventional formulations are not even plausible proxies for welfare.

In comparing the information requirements under the two approaches, it may initially appear that the preponderance rule is less demanding because it depends on fewer components. However, this impression is misleading; in fact, the underlying information requirements are quite similar. To see this point, consider the pertinent factors.

Using Figure 6, for the optimal evidence threshold, begin with the top and middle elements, the increases in expected sanctions due to a reduction in the evidence threshold and the concentrations of marginal acts. Recall from Section I.B that the reason these elements appear is that their product determines the number of acts deterred (for the left box) and the number of acts chilled (for the right box) when the threshold is changed. These products, therefore, indicate the changes in the frequencies that harmful and benign acts are committed. Multiplying these by the likelihood that acts are identified at stage one (itself a component of the expected sanctions) yields the change in the frequencies with which harmful and benign acts come before the tribunal. But it is apparent from Figure 5 that the levels of these frequencies are a central input to determining the evidence threshold for the preponderance rule and other likelihood-based rules.[87] The other elements for the preponderance rule,

87. If one was already employing some evidence threshold x^T and, moreover, it so happened that the threshold was one that satisfied the preponderance rule (or some other ex post likelihood target), one might be able to verify this by looking at the existing frequencies for the two types of acts coming before the tribunal (if these could be directly observed) and the distributions of evidence associated with each type of act. However, if not already at such a point (whether by chance or by prior calculation), one would need to know the sorts of information described in the text to determine how to set the threshold so as to meet the

the probabilities of evidence at the threshold, are, as explained in the preceding Subsection, components of the increases in expected sanctions, the top elements in Figure 6.

Next, consider the private benefits of the two types of acts, in the bottom elements in Figure 6. These do not appear directly in Figure 5, but the underlying determinants are employed indirectly. Recall that these benefits refer to those for marginal acts, and these in turn (as explained in Section I.B) will equal the expected sanctions for the respective types of acts. (Individuals on the brink of being deterred or chilled are those whose private benefits from their act just equal the pertinent expected sanction.) And these expected sanctions determine how many harmful acts are deterred and benign acts are chilled, with the undeterred and unchilled acts being the frequency with which each type of act is committed, a component of the top elements in Figure 5 (which factor is multiplied by the stage-one identification rates).

At this point, we can see that all the information needed to determine the optimal evidence threshold—except for the magnitude of social harm—is likewise necessary for application of the preponderance rule or similar likelihood-based rules.[88] Reflection on the foregoing suggests that the converse is true as well. Hence, the only fundamental informational difference between the two contrasting formulations for the evidence threshold concerns the degree of social harm from harmful acts. But presumably the legal system has pertinent information on this element (or what it considers to be a workable estimate) because it is a central determinant of the magnitude of the sanction.[89]

target likelihood. Moreover, in light of the problems identified in Subsection II.C.1, it may not be easy to iterate to an evidence threshold producing the requisite ex post likelihood, in part because it would not be obvious in which direction to move the threshold if not already at the desired target.

88. Although the text explains the underlying connections between the information requirements for each approach, they are rather subtle, and one may not be entirely convinced by such an informal exposition. For a formal treatment, see Kaplow, Optimal Burden of Proof, *supra* note 16. As explained there, the frequency distributions required for determining the ex post likelihood are the cumulative distribution functions for the benefits from the two types of acts, whereas the concentrations required for determining the optimal evidence thresholds are the corresponding density functions. (Because the density functions are simply the derivatives of the distribution functions and the latter are the integrals of the former, knowledge of one is equivalent to knowledge of the other.) Moreover, one needs to know where to evaluate each of these functions for a given evidence threshold, and those points are given by the benefits of the marginal individual of each type, which in turn constitutes the information on private benefits in the lower elements of Figure 6.

89. In a narrowly retributive system, in which the magnitude of the punishment is supposed to equal the gravity of the crime (or some given proportion thereof), this would necessarily be true. More broadly, when sanctions are set with welfare at all in mind, this information will

For civil liability, ordinary compensatory damages equal harm, which consequently must be measured.

For the most part, then, the great gulf between the formulation for the optimal evidence threshold and that for the threshold that implements the preponderance rule (or other likelihood-based rules) does not involve what information is required. Instead, the difference is in how that information is utilized. Informational requirements are a significant challenge under conventional conceptions and under welfare maximization. Regarding the former, the subtle, demanding informational basis for application of the preponderance rule seems not to have been fully appreciated.[90] In any event, even if the preponderance rule and others like it were simpler to apply, this facility would hardly constitute a sound argument in their favor.

C. Elaboration

Section II.B raises serious questions about the normative basis for the preponderance of the evidence rule and for other conventional burden of proof standards that determine liability based on a target ex post likelihood. As explained, the main factors relevant to social welfare are largely irrelevant under all such rules. Not only do conventional formulations answer a question that turns out to have almost nothing to do with what matters for society, but they also have additional features that are, on reflection, quite surprising, as elaborated in Subsection II.C.1. Subsection II.C.2 examines alternative objectives advanced in some of the literature and explains how similar problems as well as additional difficulties arise.

1. Relationship Between Level of Evidence Threshold and Ex Post Likelihood

The last topic in Subsection II.B.1 concerns how any relationship between the optimal evidence threshold and the ex post likelihood could prevail. Another basic question concerns the functional relationship between the

be important (or at least indirectly so, in determining enforcement effort and thus the expected sanction; *see infra* Part III).

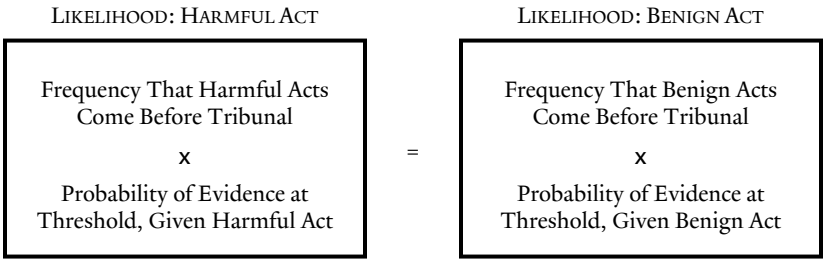
90. To implement a preponderance (or similar type of) rule, one method would be for the system designers to determine (in various classes of cases) the appropriate evidence threshold and then communicate this threshold to factfinders, just as one might imagine being done if the optimal evidence threshold were employed. *See infra* Section II.D. If instead the matter is left to factfinders in individual cases, application of the preponderance rule would be much more complicated, as the analysis in Subsection II.C.1 indicates. Additional factors influencing the selection of cases in stage one, some explored in Section IV.D, make the analysis even more challenging—see especially note 218.

evidence threshold (optimal or not) and the ex post likelihood that the individual before the tribunal committed a harmful act. It is generally supposed and seems intuitively correct that raising the evidence threshold will increase this likelihood because a higher threshold demands stronger evidence in order to assign liability. This presumed relationship, however, need not hold; it is entirely possible, and also plausible, that a higher threshold will sometimes result in a lower ex post likelihood rather than a higher one for cases that just meet the new, heightened threshold. In other words, insisting on stronger evidence may have the result that, in cases just meeting the stricter threshold, a greater fraction involves individuals who committed benign acts and are thus being sanctioned by mistake. Moreover, this fact and the analysis that underlies it have additional, unappreciated consequences.

This counterintuitive result can be understood by reexamining Figure 5, reproduced here for convenience.

Figure 5.

EVIDENCE THRESHOLD FOR THE PREPONDERANCE RULE



In particular, contemplate how the values of the products in the left and right boxes change as one raises the evidence threshold slightly. Since the left box gives the likelihood that acts generating evidence just at the threshold are harmful ones and the right box gives the likelihood that they are benign, we wish to know how the relative magnitudes in the two boxes change as the evidence threshold is increased. There are two types of effects, corresponding to the two types of elements in the boxes.

Begin with the lower elements, referring to the probability of evidence at the threshold for the two types of acts. More precisely, hold constant for the moment the frequency with which harmful and benign acts come before the tribunal, the upper elements in Figure 5. In that case, we wish to know whether raising the evidence threshold raises the ratio of the likelihood that harmful acts generate evidence just at the threshold to the likelihood that benign acts generate evidence just at the threshold. This ratio will indeed increase under

ordinary understandings of what it means for evidence to be stronger (as elaborated in the margin[91]); hence, if this were the only effect, the standard view would be correct.

There is, however, another effect, corresponding to the influence of the top elements in the boxes in Figure 5, the frequencies with which the two types of acts come before the tribunal. Since we are determining the effects of an increase in the evidence threshold, we know that the result will be to reduce the likelihood of applying sanctions to both types of acts, which implies that both deterrence and chilling will fall. In other words, more individuals will commit both harmful acts and benign acts, so more of both types of acts will come before the tribunal. The former effect further increases the likelihood that acts before the tribunal with evidence just at the (new, heightened) threshold are harmful ones, but the latter effect reduces this likelihood. Therefore, if the heightened evidence threshold reduces the chilling of benign acts relatively more than it reduces the deterrence of harmful acts, the shift in the flow of cases will be such that a higher fraction of cases entering the system will involve benign acts. Holding constant the rates of finding liability for each type of act (related to the lower elements in the boxes in Figure 5), this phenomenon would increase the likelihood that individuals found liable would be ones who committed benign acts rather than harmful ones.

Although there is certainly no guarantee that this relationship regarding the upper elements will hold, it is entirely plausible. That is, nothing in the setup of the problem rules it out or makes it somehow unlikely. Note further that this property actually tends to make raising the evidence threshold attractive. After all, it indicates that the increase does less to undermine deterrence than to alleviate the chilling of benign acts. Accordingly, it makes sense to contemplate seriously cases in which this effect is in the direction of raising the fraction of benign acts coming before the tribunal, even though this possibility is not the only one.

91. Specifically, this property corresponds to the curve relating the two probabilities of liability (which, recall, differ from the probabilities of evidence at the threshold) in Figure 3 having the shape depicted there. See the discussion of this figure at the conclusion of Subsection I.A.2. The relationship just described in the text can also be seen from the curves in Figure 1, which indicate the pertinent probabilities of evidence for the two types of acts. For example, toward the middle of that figure (between the two peaks), the probability of evidence at any value of x is rising for harmful acts but falling for benign ones, so it is apparent that the ratio increases with the threshold. Toward the right of the figure (right of the peak for harmful acts), both relative increases are negative, so for the ratio to rise, it must be true that the percentage rate of fall is lower for harmful acts (which is not obvious by visual inspection but holds true under the conditions previously described).

Let us now combine these two effects, corresponding to the two types of elements in Figure 5. The more familiar one, corresponding to the lower elements, raises the relative likelihood that an act generating evidence just at the (new, heightened) threshold is of the harmful type, but the other effect, corresponding to the upper elements, may well (but need not) lower this likelihood. If this latter effect is in the opposing direction and happens to have a greater magnitude than the former effect, then indeed the combined result of a heightened threshold would be to reduce (not increase) the ex post likelihood that an individual with evidence just meeting the higher threshold would be one who committed a harmful act. Put another way, findings of liability in cases that just pass the higher threshold may involve a higher (not lower) fraction in which sanctions are imposed by mistake.

This outcome is not a mere theoretical possibility but rather is entirely conceivable, it being an empirical matter how often and under what circumstances it will arise.[92] To illustrate this situation, suppose that there is a relatively large set of opportunities to commit benign acts, but most are just barely chilled under the current regime. In that setting, an increase in the evidence threshold may, through the reduced chilling effect, cause a relatively large increase in the flow of benign acts into adjudication, causing a notable rise in the relative rate of mistaken imposition of sanctions despite the fact that such is less likely on a per-case basis.

The foregoing analysis demonstrates that a higher evidence threshold may be associated with a lower de facto burden of proof if the proof burden is understood as a minimum ex post likelihood that cases with evidence just at the threshold involve individuals who actually are ones who committed harmful acts. Conversely, a system designer hoping to raise the burden of proof, thus defined, may in some instances need to lower the minimum required level of evidence, as paradoxical as that may seem. These consequences, two sides of the same coin, are not only possible but do not require far-fetched assumptions in order to occur.

This possibility has additional implications. Notably, it may be that a given target ex post likelihood—such as the 50% threshold of the preponderance rule—would be associated with multiple, distinct evidence thresholds. To see

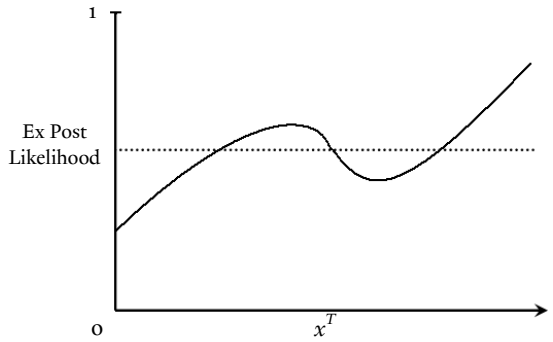
92. Note that, although an argument was given about the nature of stronger evidence implying that the former effect is in the normally supposed direction, there was no particular basis for assuming anything about the magnitude of this effect, specifically, relative to the behavioral effect. It can be shown that the ex post evidentiary effect depends on the comparison of the slopes of the two curves in Figure 1, each relative to their height, at the evidence threshold x^T in question, but the behavioral effect does not. By contrast, the latter depends on a number of factors that have no influence on the former.

this, begin with the standard supposition that, as one raises the evidence threshold, the ex post likelihood rises. In that case, when the evidence threshold is very low, we might suppose that the resulting ex post likelihood is below 50%. As the threshold is increased more and more, the ex post likelihood continues to rise and eventually crosses 50%. Once it does, it rises ever further as the threshold is increased additionally. Hence, there will be a unique evidence threshold at which the ex post likelihood equals 50%.

Now introduce the just-established possibility that the ex post likelihood falls with the evidence threshold in some circumstances. Then, after the likelihood passes 50% in the foregoing illustration, it may be that the likelihood at some point begins to fall; and, with further increases in the threshold, it may fall even more, perhaps again crossing 50%, this time moving in the downward direction. In that case, there would be two different evidence thresholds that are each associated with a 50% ex post likelihood. Moreover, if after further increases in the threshold, the ex post likelihood once again started to rise, it could now cross 50% a third time; and so forth.[93] In all, there could be one, two, three, or any number of evidence thresholds that yield a 50% probability—or a 75% probability, a 95% probability, or any other.

This possibility is obviously disturbing for a system designer trying to implement, say, the preponderance rule. It can also be problematic for a

93. Consider the following depiction:



In this figure, if the dotted line indicating the target ex post likelihood were, say, at 50%, then there would be three levels of x^T at which the rule would be satisfied. As drawn, this would also be true at targets of 49% or 51%, but not, say (interpreting the figure roughly), at 35% or 65%. Also, as drawn, for sufficiently low or high targets, say 10% or 90%, there is no value of x^T that works, which supports the next point in the text. Similarly, for the target of 50%, all levels of x^T would be insufficient if the curve happened to be much lower (everywhere below the dotted line), and all levels would be excessive if the curve was much higher (everywhere above the dotted line).

factfinder because it might be difficult to act faithfully on the command to find liability if and only if the preponderance rule—or any other rule that is based on the ex post likelihood—is satisfied in cases in which the strength of the evidence lies above the lowest threshold that works but below the highest such threshold.[94] Relatedly, in such situations the evidence threshold could be at a level higher than one of those that just hits the target likelihood and yet, when evidence is at that threshold, it would not be sufficient to meet the target ex post likelihood; and evidence just at a weaker threshold may be stronger than necessary to meet the target likelihood.[95]

A less subtle but related point is that there is no guarantee that there exists any strength of evidence such that evidence of that strength is associated with the target ex post likelihood. Taking again, for concreteness, the preponderance rule, it is possible that all levels of evidence, no matter how strong, would be below the threshold. And it is also possible that all levels of evidence, no matter how weak, would be above the threshold.[96] These scenarios are likewise attributable to the top elements in the boxes in Figure 5. For example, if pertinent opportunities overwhelmingly involve benign acts, then even fairly strong evidence may be more likely to have arisen as a consequence of a benign act, one of the many that end up before the tribunal, than a harmful act, one of the very few that were committed. Employing an ever higher threshold would increase the flow of harmful acts coming before the tribunal, as explained above, but it would also raise the number of benign acts. In the reverse case (nearly all opportunities involve harmful acts), even weak evidence may almost always be attributable to a harmful act because nearly all acts that make it to the tribunal are harmful ones. Reducing the

94. The difficulties raised in the text need to be understood as equilibrium phenomena. When there are multiple thresholds that meet a target likelihood, we need to think about one at a time. Given that one of those thresholds is taken to be chosen, it is supposed that individual actors anticipate that choice and will behave ex ante accordingly. And, if that indeed occurs, then when factfinders apply that threshold in adjudication, the stated ex post likelihood will result for cases with evidence just at that threshold. Had one of the other thresholds been chosen and anticipated by individuals ex ante, then different behavior would have resulted, producing a different flow of cases into the system, and with that different mix, it would again be true that, for cases just at the alternative threshold, the target likelihood would be produced. Put another way, the ex post likelihoods associated with a given x^T in the figure in note 93 are determined under the assumption that, when individuals decide whether to commit the two types of acts, they are supposing that the threshold that will be applied in adjudication is indeed x^T.

95. In the figure in note 93, evidence thresholds between the middle and right intersections are associated with ex post likelihoods below the target level, and thresholds between the left and middle intersections are associated with ex post likelihoods above the target.

96. *See supra* note 93.

evidence threshold ever further will, to be sure, augment deterrence, but it will also increase chilling, and there is no guarantee that the ratio of harmful acts coming into the system will not continue to be high enough to offset the weakness of the evidence.

Combining these points, we find that there may be no evidence threshold associated with a target ex post likelihood (either because all thresholds are too high or all are too low), or one threshold (as ordinarily assumed), or multiple thresholds.[97] Which situation will hold depends on the relative magnitudes of the elements in the two boxes in Figure 5 and how they change with the evidence threshold. Ultimately, what state of affairs prevails is an empirical question. The answer depends on the nature of the evidence associated with each of the two types of acts (the bottom elements in the boxes) and on the frequencies of the two types of acts that come before the tribunal (the top elements). And the latter frequencies, in turn, depend on the opportunities for the two types of acts (including how the benefits thereof are distributed) and also on the levels of expected sanctions for the two types of acts (which depend, in turn, on the nature of evidence, the evidence threshold, and also on the level of the sanction and on the process by which a portion of each of the two types of acts is selected to come before the tribunal). Given the multiplicity of determinants and the complexity of their interrelationships, it is difficult a priori to say how often the situation with one and only one threshold is likely to prevail.

All of the foregoing results are attributable to the fact that individuals' behavior—decisions whether to commit harmful acts and benign ones—is to a degree influenced by the legal system; that is, deterrence and chilling are consequences of the expected sanctions for the two types of acts, and these expected sanctions, in turn, are influenced by the evidence threshold. Understandings of conventional burden of proof conceptions, by contrast, have an ex post orientation and thus tend to neglect these phenomena. This omission is hardly inherent; indeed, proper application of the formulation captured by Figure 5 indicates—as explained, through the upper elements—that behavioral effects are actually central to the determination of ex post

97. Observe that these complications have no analogue with regard to the optimal evidence threshold. The analysis thereof, summarized in Figures 4 and 6, indicates which threshold is best with regard to effects on welfare. Once such a threshold is chosen, that threshold can itself constitute the applicable rule. Note further that, given an optimal threshold x^T, one can compute the likelihoods in the two boxes in Figure 5 and determine the ex post likelihood that an act at that threshold is a harmful act when the strength of the evidence equals x^T. However, if the rule was articulated in terms of that resulting likelihood rather than the particular x^T, there could arise the problem that other (suboptimal) levels of x^T would be consistent with the same likelihood.

likelihoods, and this accounts for the properties presented in this Subsection. However, it seems that commentators and courts' ex post perspective has tended to lead them to overlook feedbacks due to effects on ex ante behavior.

The current state of affairs is, on reflection, perplexing. It suggests that, despite the longstanding orientation toward proof burdens conceived in terms of ex post likelihoods and the substantial literature debating various implications and applications, no one has really focused on how one would implement such a proof burden in a basic setting in which individuals' behavior is influenced by legal rules—that is, when deterrence and chilling effects are present. When we attempt to take these conventional formulations seriously, which is the purpose of this Part of the Article, we find that there are fundamental problems that have escaped notice. Moreover, these problems are entirely internal, having to do with the feasibility of implementation. They stand independently of the core defect, identified in Section II.B, that conventional conceptions omit virtually all factors relevant to setting evidence thresholds optimally, that is, to maximize social welfare, and instead are based on factors that have no direct bearing on welfare. This key reminder is convenient in the sense that the problems identified in this Subsection do not need to be addressed if one actually is concerned with welfare because, in that event, Figure 5 is irrelevant in setting the evidence threshold.[98]

A further observation may be made about the relationship between the present analysis and some of the current literature: it might seem fortuitous that the formulation in Figure 5, with its troublesome upper elements, can properly be jettisoned, because the relevance of those frequencies under traditional burden of proof notions has proved controversial. A number of commentators seem bothered by the idea that the outcome in a legal case may depend on these Bayesian priors, which, as explained in Section II.A, refer to the relative frequency with which individuals who committed harmful acts come before the tribunal, with that frequency assessed before examining the evidence presented.[99] The claim is that factfinders should decide cases based on

98. It will be explained below, especially in Section IV.C, that the formulation in Figure 5 is more nearly relevant in other settings, but those will be ones in which the conundrums elaborated here do not arise.

99. *See generally* Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. REV. 439 (1986) (commenting on the debate about the role of Bayesian modes of inference in factfinding); Daniel Shaviro, *Statistical Probability Evidence and the Appearance of Justice*, 103 HARV. L. REV. 530 (1989) (criticizing skepticism about Bayesian reasoning); Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971) (criticizing the use of probability and statistics in trials, including reliance on Bayesian inference). Regarding the belief of some academics that factfinders cannot be expected to give consideration to Bayesian priors, consider Lord Hoffman's contrary

the evidence and solely on the evidence, being uninfluenced by any thoughts about the chances that the case they hear will involve an individual who in fact committed a harmful act rather than one who committed a benign act.[100]

Note initially that this objection is more an assertion than an argument, for it presumes that ignoring priors is desirable, which some simply take to be self-evident. Against this, one might consider an opposite view that may seem more compelling. If, for example, we lived in a society where it was known that the authorities often attempted to prosecute innocent individuals, perhaps factfinders should be more reluctant to convict. In the civil context, if it was understood that frivolous suits were quite frequent in a given domain, caution might again seem in order. By contrast, if cases of some type are ones that

statement in *Secretary of State for the Home Department v. Rehman*, [2001] UKHL 47, [2003] 1 A.C. 153 (H.L.) [55] (appeal taken from Special Immigration Appeals Comm'n):

> But . . . some things are inherently more likely than others. It would need more cogent evidence to satisfy one that the creature seen walking in Regent's Park was more likely than not to have been a lioness than to be satisfied to the same standard of probability that it was an Alsatian [German shepherd]. On this basis, cogent evidence is generally required to satisfy a civil tribunal that a person has been fraudulent or behaved in some other reprehensible manner. But the question is always whether the tribunal thinks it more probable than not.

100. There is an interesting distinction being made in such discussions, for the fact of the individual appearing before the tribunal, standing accused of the offense charged—or of the civil wrong alleged by the plaintiff—is itself not taken to be part of the set of evidence the factfinder should consider. Furthermore, although some object to Bayesian reasoning with regard to the fact of an individual being before the tribunal, fewer would object in principle to its application in assessing the evidence, which seems unavoidable. For example, if it is alleged that a wealthy individual, not in financial trouble, executed an elaborate plot (planning, accomplices, secret compartments in the getaway vehicle) to steal one dollar, a factfinder would find the story incredible and would demand unusually strong proof, whereas if an individual who has been fired and faced financial hardship stood accused of taking cash from the register on the way out the door, a higher a priori plausibility would be assumed. These Bayesian priors would not, in general, be based on other evidence, such as expert testimony on the background frequencies of these distinct acts, but rather on factfinders' general understandings of human nature and other features of the social system. In addition, it can be hard to separate priors about evidence from priors about who appears before the tribunal. For example, if a case is quite far-fetched, should the prior nevertheless be set implicitly at fifty percent, so that evidence slightly more likely to arise when the individual is in fact one who committed a harmful act is sufficient for liability? Or, instead, should the factfinder draw negative inferences from the lack of evidence or weakness of evidence that fails to counteract the initial implausibility of the scenario alleged to give rise to liability? (In this instance, application of an ex post likelihood standard, including the frequencies that together constitute the Bayesian prior, may serve implicitly as a production burden. The requirement of proof beyond a reasonable doubt can be understood in part in this fashion, placing a high production burden on the prosecution, one that allows a defendant to prevail without offering any evidence whatsoever if the prosecution's case is not sufficiently strong.)

prosecutors might be most reluctant to bring absent high confidence and are known to be difficult to prove, one might suppose that a decisionmaker should be somewhat more forgiving. Of course, all of these statements should be qualified (and, in a sense, discarded) because determination of the optimal evidence threshold does not in the present setting depend on the frequencies that harmful and benign acts come before the tribunal.[101] Moreover, application of a given evidence threshold x^T requires determining whether the strength of evidence x lies above x^T, which does not require ascertaining the Bayesian prior.

Some have suggested in particular that Bayesian priors be ignored in applying burdens of proof—whether the preponderance rule, the requirement of proof beyond a reasonable doubt, or other standards. In one sense, the suggestion is obscure: how can one insist simultaneously on applying a formula and on ignoring some of its elements? It is as if one was asked to choose the rectangle with the greater area, but in so doing to ignore the length of the rectangles under consideration. What seems to be meant, and is sometimes stated explicitly, is that factfinders should decide as if the ignored components were equal. For the rectangles, choosing the one with the greater area is converted into choosing the one with the greater width, even if examination of the lengths would reveal that it has less area. For the burden of proof, this transformation involves deciding based solely on the probability of the evidence for each type of act, disregarding the frequencies. Therefore, one might find liability even when, if frequencies were considered, the individual's act being of the harmful type was most improbable—which would arise when nearly all acts in the system were benign. And one would exonerate even when the frequencies implied that the case at hand almost certainly involved a harmful act—which would arise when few benign acts entered the system. A further implication is that liability would be assessed in cases where the likelihood that there was indeed a harmful act was far lower than in other cases where individuals were exonerated.

Two features of this approach may be noted. First, it really is no longer apt to refer to the burden of proof as a preponderance rule, a requirement of proof beyond a reasonable doubt, or any other formulation that suggests that the criterion involves ex post likelihoods. After all, it does not. Different language, including different jury instructions, and different rationales would be required because the transformed rules are qualitatively different from the corresponding original ones. Second, since the ex post formulations captured

101. As explained in the comparison in Section II.B, factors that determine the flows of cases are components of determinants of deterrence and chilling effects, which are relevant to welfare.

in Figure 5 have little relationship to the formulation for an optimal evidence threshold, depicted in Figures 4 and 6, one might simply return to the point that this debate should be viewed as moot, at least if the concern is with the promotion of social welfare.[102] In any event, this issue will be revisited in Section II.D, which raises the question of the extent to which, in practice, factfinders actually follow official formulations of proof burdens.

2. Alternative Ex Post Objectives

In setting a burden of proof or more broadly in judging the success of a system of adjudication in terms of error rates, commentators advance or implicitly employ a number of objectives that, like the preponderance rule and related formulations, take an ex post view of the outputs of the legal system.[103] These alternative criteria are deficient for similar reasons: their determinants have little relationship to the factors pertinent to the advancement of social

102. In some of the economics literature on burdens of proof, a decision rule that omits priors (or, equivalently, takes them to be equal) is sometimes referred to as a preponderance of the evidence rule. *See* Dominique Demougin & Claude Fluet, *Preponderance of Evidence*, 50 EUR. ECON. REV. 963 (2006); Dominique Demougin & Claude Fluet, *Rules of Proof, Courts, and Incentives*, 39 RAND J. ECON. 20 (2008) [hereinafter Demougin & Fluet, *Rules of Proof*]; Henrik Lando, *When is the Preponderance of the Evidence Standard Optimal?*, 27 GENEVA PAPERS ON RISK & INS.—ISSUES & PRAC. 602 (2002). This decision rule, which equates the values of the lower elements in Figure 5, turns out to be optimal in a very special case, one in which there are no chilling effects and the only adverse consequence of mistaken liability is to reduce deterrence of harmful acts (due to the fact that abstention has become less attractive). *See* Kaplow, *Proof Burdens*, *supra* note 16. Similar models are also used in Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307, 345-62 (1994); Louis Kaplow & Steven Shavell, *Accuracy in the Determination of Liability*, 37 J.L. & ECON. 1 (1994); and A. Mitchell Polinsky & Steven Shavell, *The Theory of Public Enforcement of Law*, *in* 1 HANDBOOK OF LAW AND ECONOMICS 403, 427-29 (A. Mitchell Polinsky & Steven Shavell eds., 2007).

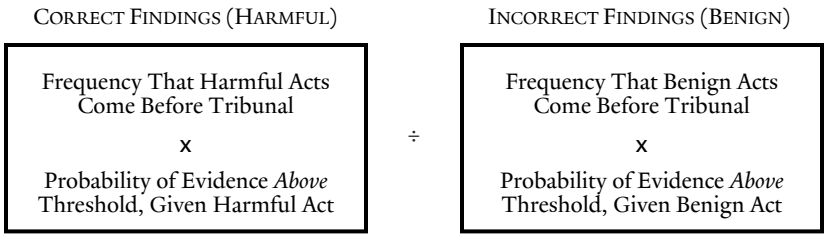
103. *See, e.g.*, Brian Forst, *Toward an Understanding of the Effect of Changes in Standards of Proof on Errors of Justice*, 41 JURIMETRICS J. 489 (2001); D.H. Kaye, *The Error of Equal Error Rates*, 1 LAW, PROBABILITY & RISK 3 (2002); Ronald J. Allen, *The Error of Expected Loss Minimization*, 2 LAW, PROBABILITY & RISK 1 (2003) (reply to Kaye); D.H. Kaye, *Two Theories of the Civil Burden of Persuasion*, 2 LAW, PROBABILITY & RISK 9 (2003) (response to Allen). It is familiar in economic analysis of the law, and more widely, that looking at outcomes of the trial or appellate process can be misleading due to selection effects. For example, if most valid cases settle or result in plea bargains, then cases going to trial will be an unrepresentative sample of all cases entering the system. By contrast, the present analysis demonstrates the folly of looking at ex post results in a setting in which there are no such selection effects; hence, the analysis reveals a deeper problem with the ex post perspective. *See generally* KAPLOW & SHAVELL, *supra* note 2, at 327-28, 372-73 (discussing limitations imposed by the ex post character of most fairness-based analysis of law enforcement).

welfare, and the formulas can have rather bizarre implications regarding what counts as an effective legal system—the latter being an implication of the former. This Subsection considers a number of such objectives, and it will be apparent that others based on similar considerations would fare no better.[104]

First, consider judging a system based on the ratio of correct to incorrect assignments of liability—true to false convictions in the criminal context. (Examination of other ratios, like that of correct to total assignments of liability, or of incorrect to correct or total assignments of liability, would lead to essentially the same conclusions.[105]) This fraction is depicted in Figure 7.

Figure 7.

RATIO OF CORRECT TO INCORRECT FINDINGS OF LIABILITY



The upper elements in the two boxes are, of course, the same as the upper elements in the two boxes in Figure 5, for the preponderance rule. Once again we have factors that differ from (do not appear in) the formulation for the optimal evidence threshold, Figures 4 and 6.

The lower elements in Figure 7 importantly differ from those in Figure 5, as indicated by the italics: here, we are considering the probability that evidence is above rather than at the threshold. The reason is that, for this ratio of correct to mistaken findings of liability, we are interested in knowing the total number of

104. For prior discussion of many such criteria, with an emphasis on their divergence from the dictates of probabilistic standards of proof, see Michael L. DeKay, *The Difference Between Blackstone-Like Error Ratios and Probabilistic Standards of Proof*, 21 LAW & SOC. INQUIRY 95 (1996). DeKay further argues that we should care about effects on utility rather than arbitrary probability ratios, although he implicitly takes behavior to be exogenous and thus provides arguments more pertinent to settings akin to that analyzed in Section IV.C on the regulation of future conduct.

105. To see that the analysis is virtually identical, let C denote correct findings of liability, I incorrect findings, and $R = C/I$. Simple algebraic manipulation confirms that $C/(C + I) = R/(1 + R)$; $I/C = 1/R$; and $I/(C + I) = 1/(1 + R)$. Therefore, the determinants of R are also the determinants of each of these other ratios, and the first of these three alternatives is always rising in R while the second and third are always falling in R, so properties of R can readily be translated into those of any of the other three ratios (and conversely).

each type of finding—rather than, with the preponderance rule, knowing the characteristics of cases just at the threshold. To determine, say, the number of correct findings, we need to know how many cases actually involving harmful acts come before the tribunal—the upper element in the left box in Figure 7—and what portion of those cases result in a finding of liability. The latter will be all those cases with evidence stronger than the minimum requirement, that is, above the threshold. The analysis for benign acts is the same.

To see this difference in another way, consider Figure 1 and take, for example, a very low threshold (toward the left of the figure). In that instance, the probability of evidence being just at the threshold is low for both types of acts, but the probability of evidence being above the threshold (the area to the right) is high for both. For a threshold in the middle of the figure, say at the crossing point of the two probability distributions, the likelihood at the threshold is the same for both types of acts, but the probability of evidence above the threshold (the area to the right) is still very large for harmful acts but is fairly small for benign acts. The probability of evidence above the threshold is simply the probability of liability for a given threshold, which is depicted in Figure 2. Thus, one way to see the difference between relative probabilities of being at versus above the thresholds is to compare Figures 1 and 2 for any given evidence threshold.

The probabilities of liability for harmful and benign acts, the lower elements in the boxes in Figure 7, also do not appear in Figures 4 and 6, for the optimal evidence threshold—whereas with the preponderance rule, Figure 5, the lower elements are a component of the top elements in Figures 4 and 6, as explained in Section II.B. Accordingly, taking this ratio of correct to incorrect findings of liability as a criterion for system quality is even more removed from the formulation of the optimal evidence threshold than was the preponderance rule. We now have no elements in common.[106]

106. Similar analysis would be applicable under the criterion that the number of the two types of error—false exonerations and mistaken findings of liability—should be equated. *See, e.g.*, FINKELSTEIN, *supra* note 3, at 65-69 (suggesting that equalizing errors between the parties in civil cases is an independent goal, which must be balanced against the competing, strong policy of reducing the total number of errors). That formulation can be represented by making two alterations in Figure 7: substitute an equals sign for the divided by sign, and, for the lower element in the left box (only), substitute "below" for "above." The latter implies that the product in the left box becomes the number of cases involving harmful acts that result in exoneration (rather than liability), and the criterion calls for this number to be equated to the number already represented in the right box. This formulation, like the ratio in the text, has no elements in common with that for the optimal evidence threshold.

Consider also a criterion requiring an equal chance of error for cases coming before the tribunal. This criterion would be the same as that just considered, except ignoring the upper

The lesson is that, if we care about social welfare, this sort of ratio tells us nothing about whether a system is functioning well. To illustrate this point, consider first a case of an extremely well functioning legal system, one that deters virtually all harmful acts and chills almost none of a huge number of benign acts. In that case, even if adjudication is reasonably accurate, we know that the value in the left box has to be close to zero (because the upper element is near zero and the lower element cannot exceed one). And the value in the right box may well be notably above zero (the upper element is taken to be quite large, so even if the second element is quite small, say, only a few percent of cases with benign acts result in liability, the product will be nontrivial, that is, not close to zero). Accordingly, the ratio of correct to incorrect findings of liability will be near zero, which initial intuition would have interpreted as a terrible score.[107]

Next, consider a different scenario:[108] Individuals decide whether to put money in parking meters, and their only motivation for doing so is to avoid having to pay tickets for meter violations. The police give tickets entirely randomly to a portion of cars parked at meters (that is, without regard to whether money was put in the meters). Adjudication involves a coin flip, with half of these tickets being upheld. In this scenario, no one would put money in a parking meter. Hence, everyone would in fact be truly liable. Therefore, the value in the left box in Figure 7 would be large and that in the right box would be zero. The resulting ratio is infinite[109] – surely a perfect score according to the criterion of the ratio of correct to incorrect assignments of liability.

Combining these examples, we can see that horrible systems can yield a perfect score and wonderful systems a terrible score, confirming that this sort of ratio is a nonsensical way to assess the performance of a system of

elements in the boxes (the frequencies that the two types of acts come before the tribunal). Again, there is no overlap with determinants of the optimal evidence threshold.

107. An implication is that, if the legal system were to insist that this ratio exceed some minimum, it may be infeasible to maintain a well-functioning system, a point reinforced by the next example. *See also* Ezra Friedman & Abraham L. Wickelgren, *Bayesian Juries and the Limits to Deterrence*, 22 J. L. ECON. & ORG. 70 (2006) (presenting a model in which insistence on a minimum ex post probability that defendants are truly guilty places a floor on the level of crime).

108. This scenario departs in one respect from the framework presented in Subsection I.A.1 because individuals are choosing between the two types of acts, but, as suggested there, this difference is inconsequential for many purposes, including the present one.

109. Actually, the ratio, which involves dividing by zero, is undefined. But we could modify the example ever so slightly by assuming that there is a tiny probability that one person puts money in the meter. As this probability approaches zero, the expected value of the ratio approaches infinity.

adjudication, despite the initial appeal of the criterion. But on reflection, such odd implications should have been anticipated because the criterion has no relationship to the formulation for the optimal evidence threshold: that which is aimed at maximizing social welfare.

Some additional ex post objectives will now be considered more briefly.[110] Another criterion would be to minimize the number of mistaken findings of liability.[111] That number is the value in the right box in Figure 7, and we have already seen that its components have essentially no overlap with the determinants of an optimal (welfare-maximizing) evidence threshold.[112] Also,

110. No attempt is made to be exhaustive. To consider another ex post objective, suppose that the criterion is to minimize the fraction of mistakes in those cases that come before the tribunal. Again, the formulation for this criterion has no connection with that for an optimal evidence threshold. And, as a consequence, there can be anomalies. For example, deterrence will reduce the flow of harmful cases into the system, which reduces both the numerator and denominator of this fraction, and if the fraction of harmful acts resulting in liability is high, the ratio could worsen as a consequence. Analogously, greater chilling might, though need not, make the system look better. Note also that flooding the system with extremely weak cases involving almost entirely benign acts (say, the police engage in a random sweep of unlikely suspects)—nearly all of which result in ultimate exoneration—would improve this ratio, suggesting that the system is performing better. And dropping all but a tiny number of cases, leaving only sure winners (or a tiny number that are sure losers) might result in an excellent ratio but virtually eliminate deterrence.

111. If instead the criterion were to minimize the total number of mistakes, combining the two types—and if, moreover, one took behavior as given—then the preponderance rule would be indicated, a justification commonly noted in the literature. *See, e.g.*, FINKELSTEIN, *supra* note 3, at 66-67; Ball, *supra* note 7, at 822-23; James Brook, *Inevitable Errors: The Preponderance of the Evidence Standard in Civil Litigation*, 18 TULSA L.J. 79, 86 (1982); Neil Orloff & Jery Stedinger, *A Framework for Evaluating the Preponderance-of-the-Evidence Standard*, 131 U. PA. L. REV. 1159, 1161-63 (1983); Redmayne, *supra* note 7, at 169-70. Note that if one cared about the total number of mistakes in fact, and not conditional on behavior, which varies with the burden of proof, it is unlikely that the preponderance rule would minimize mistakes. Many of the comments in the text on the criterion of minimizing the number of mistaken findings of liability would be applicable to this more inclusive criterion.

112. It is interesting to reflect on why so many who view the legal system, especially for criminal cases, fixate on mistaken convictions of the innocent, as reflected in discussions of the so-called Blackstone ratio. *See, e.g.*, DeKay, *supra* note 104 (criticizing the focus); Alexander Volokh, n *Guilty Men*, 146 U. PA. L. REV. 173 (1997) (describing and documenting the practice). Laplace and others cited and quoted in note 42, by contrast, advance the more compelling view (although still incomplete in light of this Article's analysis) that society should minimize unjust harm suffered by innocents, which would focus on the sum of the false conviction rate and the victimization rate. *See also* Larry Laudan, *The Rules of Trial, Political Morality, and the Costs of Error: Or, Is Proof Beyond a Reasonable Doubt Doing More Harm than Good?*, *in* 1 OXFORD STUDIES IN PHILOSOPHY OF LAW 195, 199-200 (Leslie Green & Brian Leiter eds., 2011) (presenting data on crime, supplemented by conjectures on false conviction rates, suggesting that the likelihood of being a victim of a violent crime, such as rape or murder, is orders of magnitude higher than the likelihood of being falsely convicted

the two examples just given illustrate how problematic this criterion can be. In the first example with the well-functioning system, the number may be fairly high (because the number of unchilled benign acts is huge), and in the second example of the entirely capricious legal system, that number is zero. More broadly, one could guarantee a zero (perfect) score by elimination of the legal system, or come very close to it by screening out—say, through the exercise of prosecutorial discretion or application of an extreme summary judgment hurdle—all but a tiny number of cases, leaving only ones in which the evidence is overwhelming (or only ones with extremely weak evidence that would almost surely result in exoneration).[113]

In addition, for reasons given in Subsection II.C.1, it is not even true that the number of mistaken findings of liability necessarily falls as the evidence threshold is increased. Recall that raising the threshold reduces the chilling of benign acts, so more of them are committed; hence, more benign acts come before the tribunal. Since this effect can be larger than that from the fall in the probability of liability for a given benign act due to the higher threshold, the number of benign acts subject to sanctions can rise.[114] Put another way, a criterion of minimizing the number of mistaken findings of liability is advanced by chilling effects and undermined by their reduction: that is, from

for such a crime). A hypothesis is that standard views in part reflect framing effects: in considering the standard of proof, one focuses on a criminal trial and on the defendant who is before the tribunal, not on the statistics about the expected number of additional crimes on account of the deterrence reduction that results from a higher evidence threshold. This perspective helps to explain why prosecutors try to focus the factfinder's attention on the victim. (As a further explanation, particularly for analysts of the legal system, a strong focus on victims seems to be avoided because it is associated with a retributive urge that many scholars believe to be inappropriate. *See generally* KAPLOW & SHAVELL, *supra* note 2, at 352-71.) This strong aversion toward mistaken findings of liability seems far weaker or even absent in the civil context, where in the United States a preponderance of the evidence suffices. In such cases, the victim is before the court—as the plaintiff—so either outcome, if mistaken, harms an innocent individual who is directly before the factfinder and salient in the mind of an analyst who is contemplating a civil trial.

113. Similar conclusions follow if the objective, in whole or in part, is to minimize the number of mistaken exonerations. For example, under the approach in Thomas J. Miceli, *Optimal Criminal Procedure: Fairness and Deterrence*, 11 INT'L REV. L. & ECON. 3 (1991), it would improve fairness in this sense if a large number of probably guilty individuals were allowed to go free without trial rather than bringing them into adjudication, knowing that in all likelihood some of them would be mistakenly set free. Focusing on outcomes in adjudication rather than viewing all outcomes can be quite misleading, as discussed in KAPLOW & SHAVELL, *supra* note 2, at 291-378.

114. As explained previously, it can also rise relatively more than does the number of harmful acts resulting in sanctions, so the ratio discussed just above could fall rather than rise as the evidence threshold is increased.

this perspective, the more the legal system discourages benign behavior, the better (all else equal).

Finally, consider a criterion under which the likelihood of mistakes is supposed to be equal for those committing the two types of acts.[115] For harmful acts, mistakes occur when individuals are not identified at stage one and when, although identified at stage one, they are exonerated in adjudication. Put another way, mistakes occur except when such individuals are both identified and found liable. For benign acts, mistakes require the conjunction of identification at stage one and mistaken assessment of liability at stage two. Insisting that these two likelihoods be equated again involves a formulation whose elements are entirely distinct from those that determine the optimal evidence threshold.

Note further that this criterion may be impossible to meet: If less than half of harmful acts are brought into the legal system at stage one, the error rate for harmful acts as a whole necessarily exceeds 50%. Now, assuming that the rate that benign acts are identified in stage one is no higher than that for harmful acts, we could not achieve equality (which requires exceeding a 50% error rate) even if all individuals who commit benign acts and reach adjudication are found liable. Note that a higher rate of mistaken imposition of liability helps us move closer to this equality criterion, although it still falls short. What is in fact needed is rounding up additional individuals who committed benign acts so that more can be subject to adjudication and found liable. The superficial appeal of this equality criterion rapidly dissolves upon examination; it too can be perverse.

D. Comparison Revisited: Theory Versus Practice

The foregoing analysis takes conventional conceptions of the burden of proof on their face. Canonical statements, discussions, and rationalizations by commentators focus on the minimal requisite ex post likelihood that the individual before the tribunal committed the harmful act. This Section considers two questions motivated by the preceding analysis in this Part. First, to what extent do factfinders actually adhere to such commands rather than implicitly putting at least some weight on the likely consequences of their determinations, perhaps along the lines of the welfare-based analysis in

115. By contrast, note 106 considers the criterion that the number of the two types of mistakes—rather than the ex ante likelihood for those committing the two types of acts—should be equal.

Part I?[116] Second, how could the burden of proof be reformulated to attend more explicitly to welfare considerations? Both inquiries are brief and speculative, involving important empirical matters about which little is known. The analysis here should be understood as posing thought experiments rather than advancing suggestions for reform.

Beginning with the first inquiry, a number of types of deviation between idealized understandings of conventional proof burdens and actual practice seem possible. Most broadly, factfinders may not understand instructions very well, or they may not take them very literally, or they may disregard them.[117] More relevant for present purposes, it is questionable whether factfinders

116. Conventionally, in discussing these issues in the United States, courts and commentators have in mind juries as factfinders, but similar analysis is applicable to judges, agencies, or other decisionmakers. On empirical questions, answers may differ substantially, although it is not obvious how great such differences would be or just how they would be manifested.

117. As one important indication that the preponderance rule is not applied literally, consider the discussion of tiebreaking in note 34. It is explained that actual, precise ties formally have a zero probability. Therefore, aggressive disputes about who wins in the event of a tie suggest that the party bearing the burden under a preponderance rule is understood to have to prove a substantially higher likelihood than fifty percent, or in any case a notably different likelihood than would be the case if the legal command broke the tie the other way. Another consideration is that the analysis of Subsection II.C.1 shows that it is possible that multiple evidence thresholds exist or that no possible evidence threshold satisfies the preponderance rule or any other particular likelihood-based rule, in which case there is no clear meaning to taking such a rule literally.

Furthermore, as noted in Subsection II.C.1, some have expressed an aversion to taking into account Bayesian priors (corresponding to the upper elements in Figure 5) on the ground that factfinders should consider only the evidence before them or in any case do not have the relevant information or the understanding to employ Bayesian analysis, which may also imply that they do not adhere to likelihood-based instructions. But it was already explained that these points are not obviously correct: individuals do to an extent think in these ways in other settings, and it seems unlikely that they would ignore, for example, whether a characterization of events proffered by a party was a priori quite unlikely or bizarre versus entirely ordinary human behavior. *See supra* note 99; *see also* Shari Seidman Diamond & Neil Vidmar, *Jury Room Ruminations on Forbidden Topics*, 87 VA. L. REV. 1857, 1860-63 (2001) (describing studies of jurors and other decisionmakers actively incorporating expectations and preconceptions); *id.* at 1866-1904 (reporting on videotaped actual jury deliberations in which many juries spontaneously discussed legally forbidden subjects of plaintiffs' insurance and fee arrangements, drawing on their own knowledge of the matters). In addition, information limitations are significant regarding the evidence itself (the lower elements in Figure 5, which assume knowledge of the two probability distributions in Figure 1), so it is hardly clear that factfinders would gravitate toward a modified likelihood-based assessment confined to these other components. *See also supra* note 100 (raising the difficulty of distinguishing between priors about acts before the tribunal and priors used for inference about evidence in the case at hand).

routinely pay no attention whatsoever to considerations of consequences;[118] after all, individuals typically do attend to consequences when making decisions in everyday life.[119]

Consider the civil context. In adjudicating a contract dispute, factfinders may well appreciate that a general failure to find a contract breach when one in fact occurred might encourage opportunism and discourage individuals from entering into sensible contracts in the first place. Likewise, they plausibly would understand that a practice of finding breach when there was none would discourage useful contracts and induce those in contracts to behave in excessively cautious ways.[120] Moreover, there are various ways that advocates might encourage such thinking, such as through subtle suggestions during opening and closing arguments, through cross-examination, or by the manner in which witnesses are prepared to answer questions at trial. Note that such appeals, upon reflection, should be understood as attempts to persuade jurors not to follow the command to find liability if and only if evidence meets, say, the more-likely-than-not standard—keeping in mind that, as explained in Subsection II.B.1, such consequentialist thinking is inapposite to assessing this likelihood.[121]

118. Some suggestive evidence on current jury decisionmaking is offered by Diamond and Vidmar. *See* Diamond & Vidmar, *supra* note 117, at 1866-1904. They report that videotaped actual juror deliberations in tort cases contained frequent references to the legally forbidden subjects of plaintiffs' insurance coverage and attorney fee arrangements, which arose in discussions of what damage awards would appropriately compensate plaintiffs. This behavior displays a concern for consequences: not for ex ante behavior, but rather for whether a victim is made whole. It is notable that jurors' attention to these consequences may have been the product of jury instructions specifically directing them to determine reasonable compensation. On one hand, this linkage may raise doubts about whether jurors would have brought in such considerations without direction; on the other hand, if the attention to consequences is attributable to their following instructions, it may be that different instructions highlighting different consequences would be effective.

119. Additionally, in many settings—such as in raising children and dealing with coworkers—individuals appreciate that many decisions and actions affect others' subsequent behavior by influencing expectations about how they will be treated. Accounting for the lesson that will be taught or the message that will be sent is a basic ingredient of common sense. As mentioned below in the text, however, it is an empirical question how much factfinders in various settings attend to consequences in general and effects on ex ante behavior in particular and how that attention might depend on the instructions they are given (or other aspects of the adjudication environment).

120. Or, in a tort case, factfinders would worry that failure to assess liability might encourage dangerous behavior whereas improperly assigning liability could discourage legitimate activity or inappropriately drive up its costs.

121. This state of affairs exists despite counsel being precluded from explicitly advocating nullification—i.e., that the factfinder ignore legal instructions. *See, e.g.*, United States v.

We might suppose that a similar situation prevails in the criminal context, where the requirement is proof beyond a reasonable doubt. Indeed, the use of the term "reasonable" can be understood as an invitation to engage in such consequentialist deliberation. In criminal cases, prosecutors may find subtle ways to advocate for conviction in part on the ground that it is important to send a message to prospective criminals, and they otherwise attempt to bring to mind concerns for victimization that would result from too strict a demand for proof.[122] Likewise, defense attorneys may attempt to generate fears in factfinders' minds about how they too might be falsely accused, and so forth.

This idea is related to some commentators' resistance to probabilistic interpretations of proof burdens on the ground that they instead should be understood as calling upon the factfinder to formulate a belief about the matter in question.[123] In practice, individuals' beliefs are associated with requisite probabilities that depend on the circumstances. Consider, for example, the likelihoods implied by the following sorts of statements: "I believe that 1 + 1 = 2." "I believe it will not rain on our picnic tomorrow." "I believe I should get a flu shot." "I believe that murder is evil." Obviously, these statements do not all correspond to a more-likely-than-not standard. Nor do they all imply that the statement is true beyond a reasonable doubt, although some do. Rather, they convey a level of confidence that is ordinarily deemed appropriate to the context. Furthermore, some of these statements, particularly those directed toward decisions (going on a picnic, getting immunized), undoubtedly reflect consequentialist analysis, at least implicitly. For example, one might think that

Dougherty, 473 F.2d 1113, 1130-37 (D.C. Cir. 1972) (tracing the origins of the current approach to United States v. Battiste, 24 F. Cas. 1042 (C.C.D. Mass. 1835) (No. 14,545), and Sparf v. United States, 156 U.S. 51 (1895)).

122. Explicit appeals to consequences by prosecutors in criminal cases are consistently condemned by appellate courts, although they are sometimes deemed harmless error. *See, e.g.*, United States v. Roberts, 986 F.2d 1026, 1031-32 (6th Cir. 1993) (deeming as harmless error, partly in light of a jury instruction to decide the case solely on the evidence and not the lawyers' statements, a prosecutor's call on jurors to put a stop to dope dealers carrying guns); United States v. Barlin, 686 F.2d 81, 93 (2d Cir. 1982) (failing to find reversible error in the prosecutor's statement that jurors had a duty to do something about drug trafficking in their community); United States v. Barker, 553 F.2d 1013, 1025 (6th Cir. 1977) (reversing a conviction for many reasons, including a prosecutor's statement to the jury that acquittal in a bank robbery case would be tantamount to opening all the banks to make their money freely available to robbers); Brown v. United States, 370 F.2d 242, 246 (D.C. Cir. 1966) (reversing a conviction on multiple grounds, including the prosecutor's statement that failure to convict would lead to martial law).

123. *See* sources cited *supra* note 78. The occasionally used requirement of clear and convincing evidence specifically asks the factfinder to form a conviction about the matter. The same appears to be true of formulations in Continental legal systems, such as the demand for an *intime* (inner) conviction in both civil and criminal cases. *See* sources cited *supra* note 4.

a flu shot has only a ten percent probability of being useful but, in light of its low cost and potentially large benefit, one might also believe that it is almost certainly a good idea; in this case, the confidence in the belief (that I should get a flu shot) is high although the estimated probability that the shot will truly help is fairly low. The converse—a high probability of benefit, but a low level of belief, because costs are high and benefits low—is also possible.

The question of what probability factfinders actually associate with, say, the preponderance of the evidence rule—and how that minimum required probability varies by context—is an empirical one. Furthermore, it is one about which little is known. Answers to surveys on the meaning of "more likely than not" may convey little, for the suggestion here is that its meaning in practice can depend very much on the circumstances. Therefore, there probably is no one answer, maybe not even approximately,[124] and an answer given in a vacuum may not provide a reliable basis for prediction of actual factfinders' behavior. Substantial investigation is required to better understand the operational significance of existing proof requirements and, moreover, the extent to which they might implicitly depend on the sort of welfare-based reasoning presented in Part I. Perhaps more important, what matters most for deterrence and chilling effects is not what factfinders actually do but how they are perceived to decide cases, whether under existing formulations or possible alternatives.[125]

The second inquiry concerns the possibility of bringing welfare consequences more explicitly into burden of proof rules.[126] As mentioned, this discussion

124. Consider, for example, the range of answers reported by the sources cited in note 78 to the probability associated with a requirement of proof beyond a reasonable doubt. Of particular interest, Solan, *supra* note 78, at 119-33, catalogues a wide range of probabilities and conviction likelihoods associated with different forms of the "beyond a reasonable doubt" jury instruction in varying settings, some real and some hypothetical.

125. First, because most cases settle (including through plea bargaining) and it makes sense to suppose that such settlements reflect expected outcomes in adjudication, what matters for settlement is what parties, advised by their lawyers, expect factfinders to do. Moreover, since even experienced trial lawyers can have only limited insight into factfinder behavior, their advice may not mirror real decisionmaking very precisely. Second, at the time individuals choose whether to commit acts—when they may or may not be advised by lawyers (who themselves may well not be litigators)—their choices will only loosely reflect how subsequent adjudication would truly play out.

126. A different strategy would be to retain existing proof burdens and adjust substantive law by context, such as by adding elements if it is desired to make it more difficult for plaintiffs or the government to prevail. Although undoubtedly some substantive rules reflect this approach, there are disadvantages. Most directly, if the concern is, say, with excessive false positives, a better outcome is generally obtained by demanding stronger evidence than by instead insisting on evidence of something else that diverges from what is truly of social

should be understood primarily as posing thought experiments given its brevity and the paucity of empirical evidence on both optimal evidence thresholds and on the just-addressed question of how factfinders would actually respond to one or another formulation.

To begin, one could imagine a rule that spoke fairly directly in terms of consequences. In the contract breach illustration, a factfinder might be asked to come to a reasonable conclusion in light of the risks, on one hand, that failing to find a breach of contract due to an overly stringent demand for proof might encourage opportunism and discourage contracting and, on the other hand, that finding breach based on flimsy evidence might itself chill contracting and also induce excessive caution in contractual performance.[127] Many variants, more or less case-specific, might be imagined.[128] Whether such an approach constitutes good policy—in general or perhaps just in certain settings—depends on how factfinders would actually behave and on how the welfare consequences that result compare to those under other formulations, including the conventional ones.[129]

relevance. In addition, it should be kept in mind that seemingly transsubstantive proof burdens, such as the preponderance rule, do require highly context-specific information just as does a welfare-based rule, as explained in Subsection II.B.2.

127. Rigorously applying the formulations in Figure 4 or 6 would be daunting, but this does not rule out the possibility of an intuitive consideration of the likely weight of the competing factors—just as the challenge of properly applying the preponderance rule's actual formulation in Figure 5 (which, as explained in Subsection II.B.2, has information requirements similar to those of the welfare-based test) does not lead us to conclude that factfinders could not possibly do anything remotely like it.

128. A basic question about proof burdens concerns the level of generality with which they should be expressed. In the United States, a single requirement is applied in most civil cases and another, more stringent requirement, in criminal cases. But there are exceptions, such as the occasional use of a rule requiring clear and convincing evidence in civil settings and of burden-shifting or redefinition of elements in the criminal context. And greater variation can be imagined. As Section IV.C suggests, administrative agencies in many realms probably already employ context-sensitive cost-benefit analyses, and even traditional rules for standard court cases could vary much more than at present. Moreover, it should be kept in mind that even a single requirement, like preponderance of the evidence, depends on many context-specific factors; indeed, as Subsection II.B.2 explains, its underlying information requirements are actually similar to those of a case-by-case assessment of welfare effects. Additionally, the text earlier in the present Section suggests that the apparently one-size-fits-all conventional rules may already be applied by factfinders in a manner that depends importantly on the circumstances of particular cases much as would a case-specific welfare-based standard that was explicitly open-ended.

129. One might also be concerned about how different proof standards would influence the legitimacy of the system of adjudication as perceived externally. Specifically, likelihood-based criteria might, at first glance, seem more appealing because they are directed at the truth of the matter. Nevertheless, because citizens attend to consequences of important

This comparative assessment suggests an alternative approach. Suppose that commands like the conventional ones or some others could be shown to produce more predictable factfinder behavior.[130] In that event, one might choose proof burdens as follows: First, employ the sort of analysis in Part I to determine (estimate) the optimal evidence threshold in a given context. Second, ask what de facto evidence threshold would likely (perhaps on average) be employed by the pertinent type of factfinder under various verbal formulations of the proof requirement.[131] Third, pick that formulation which produces a de facto threshold that, crudely, is closest to what is optimal.[132]

For example, in some cases, a preponderance rule might be the best (albeit imperfect) choice, although not because it is the right rule in principle and not because factfinders necessarily apply it literally.[133] Rather, it may be superior because, given how factfinders would actually behave under it, they would tend to come fairly close to the optimum in most cases—or at least closer than under alternative formulations. In another setting, perhaps adjudicators would, on average, be too lax under a preponderance instruction, so employing a clear and convincing requirement or proof beyond a reasonable doubt would do better. Sometimes even those choices might be too weak, so a new, tougher

decisions in their own lives and expect their government to do likewise, they might be surprised and upset rather than comforted to learn that judgments in adjudication were based exclusively on ex post likelihoods, with factfinders being forbidden from attending to whether their decisions might support productive interactions or instead wreak havoc on society. Recall the medical decisionmaking illustration from the Introduction, where it seems difficult to imagine that a likelihood-based method that ignored consequences could arise or would command any respect, whether from medical professionals or the lay public. On the general question of how system legitimacy fits into a welfare-based approach, see KAPLOW & SHAVELL, *supra* note 2, at 286-87, 407-08, and Kaplow, *supra* note 102, at 395-96.

130. This premise is, as stated, a supposition, and one that may well be incorrect. *See, e.g.*, McCauliff, *supra* note 77, at 1324-33 (showing survey responses of federal judges that display highly varying probability estimates associated with various conventional proof burden formulations); Solan, *supra* note 78, at 119-33 (discussing the wide range of probabilities associated with the "beyond a reasonable doubt" standard when formulated in different ways and employed in different settings).

131. For example, McCauliff, *supra* note 77, at 1332 tbl.8, presents the mean percentages associated with nine verbal formulations in a survey of federal judges, with the range spanning roughly from 30% to 90%. It must be kept in mind, however, that these answers were in a vacuum and showed substantial variation for most formulations.

132. The statement in the text is intentionally rough in many respects. For example, the best choice would not necessarily be that which had an average de facto threshold closest to the optimal threshold because errors in both directions are not generally of equal weight and the magnitude of error costs in a given direction are nonlinear.

133. Indeed, its desirability may be due in significant part to their failure to do so, if the deviations tend to correlate positively with likely welfare consequences.

phrasing may be superior, perhaps a demand for almost absolute certainty. Likewise, the preponderance rule, as it would typically be employed, might be too demanding in certain types of cases; there, a weaker command, like that there exist a reasonable basis for the plaintiff's case, may perform better. In each context, the instruction would not be chosen based on whether it was correct in some intrinsic or Platonic sense—after all, unless it is couched in terms of the explicit welfare-based analysis of Part I, we know that it is not—but rather because its predicted consequences result in greater welfare than do those of the alternatives.[134]

Yet another strategy is to explore ways of formulating an instruction explicitly in terms of the evidence threshold itself. If analysis determined, say, that x^T was the optimal evidence threshold in some setting, a factfinder might be instructed to assess liability if and only if the strength of the evidence exceeds x^T. This method has the virtue that the factfinder is not asked to perform either the welfare-based calculations of Figure 4 or 6 or to undertake the Bayesian analysis of Figure 5, both of which make heavy informational demands. If the only or primary evidence is the reading on some meter—perhaps a vehicle's speed registered by radar or an individual's blood-alcohol level measured by a blood test—this method might be workable and may well be best. But in the broad range of cases, there are two complications. First, the power of evidence rarely is indicated by some instrument reading and often does not lend itself to verbal depictions that communicate very precisely. Second, many types of evidence of varying strengths, whose force depends on complex interactions, must ordinarily be aggregated by the factfinder.

In these instances, it is nevertheless possible in principle to state an evidence threshold, specifically, in terms of a likelihood ratio: the ratio of the likelihood that such evidence (taken together) would be generated by a harmful act to the likelihood that such evidence would be generated by a benign act. These likelihoods correspond to the height of the probability

134. As this discussion suggests, it would be natural to include in the initial set of alternatives the familiar rules and perhaps some others. In performing empirical work, it does seem appealing to assess as well more explicit, welfare-based formulations like that suggested previously (in the contracts illustration) and to experiment with variations. After all, it is not unreasonable to contemplate that an instruction that is explicitly in terms of the relevant considerations might perform better than instructions that are explicitly in terms of factors that we know (from Section II.B) to exclude all relevant determinants and substitute ones that are irrelevant to what really matters. The present claim is not that this point is true, but merely that it has sufficient prima facie plausibility to be taken seriously in subsequent investigations. Relatedly, there may be other tests that turn out to be best that are neither explicit, fairly direct welfare assessments nor conventional standards that are wholly divorced from consequences.

distributions in Figure 1.[135] As explained in Section I.B, these likelihoods are subcomponents of the upper elements of Figures 4 and 6. Phrased slightly differently, these likelihoods are the probabilities of evidence being at the threshold, given a harmful or benign act, respectively—that is, the lower elements in Figure 5.[136] An implication of the latter observation is that conventional rules suppose that factfinders can ascertain these values.

The feasibility and desirability of this approach is difficult to assess. The appropriate ratio depends on all the other factors in Figure 4 (or 6), so the stated ratio would ideally vary greatly from one type of case to another. Of course, under the preceding alternative that may employ more conventional formulations, the appropriate verbal test should also differ across contexts. In addition, we know little about how factfinders would actually behave under such an instruction.

Whether using an explicit, welfare-based formulation, choosing different formulations by context, stating instructions as likelihood ratios, or doing something quite different would prove to be best is an empirical and practical question about which little more than conjecture is possible at this juncture. The purpose of this analysis, which merely scratches the surface, is to sharpen the point that we need to give much more explicit attention to this basic question of system design. Perhaps we are lucky in that factfinders currently apply conventional burden of proof formulations—which, if taken literally, are fundamentally misconceived—in a manner that attends significantly to welfare-relevant consequences, and it may even be the case that society cannot do much better.[137] One reason might be that instructions about the burden of

135. *See supra* note 30 and accompanying text.

136. An implication is that, if one did want to implement the preponderance rule or other ex post likelihood-based rules but one did not want the factfinder to be concerned with the frequencies (the base rates or Bayesian priors), one could in principle employ such a likelihood-ratio instruction. Note that the instructed likelihood ratio for, say, the preponderance rule, would need to vary greatly by context, sometimes being much higher and other times much lower than one to one, as discussed in Section II.A. A related point is that these likelihoods or probabilities of evidence for the two types of acts are not the same as the probabilities that the individual before the tribunal committed one or the other type of act, precisely because the latter depends importantly on the base rates whereas the former does not.

137. One basis for such a conclusion is the fact that contracting parties do not uniformly provide for alternative dispute resolution (ADR) systems that employ alternative proof burdens. On the other hand, many factors go into the choice whether to employ ADR and how it is designed. Additionally, many forms of ADR are seen as operating informally, not necessarily basing outcomes strictly on existing legal rules, either procedural or substantive, so perhaps the frequent use of ADR does reflect a preference for a different decisionmaking criterion.

proof have little effect in any event, and factfinders instead use their common sense, which perhaps is the most that can be expected of them. But it is hard to know whether this is true, or whether we now fare quite poorly, unless we ask the relevant questions and attempt to determine the answers.[138]

III. INTERACTIONS BETWEEN THE BURDEN OF PROOF AND OTHER FEATURES OF THE LEGAL SYSTEM

This Part and Part IV return to the normative question of how to design the legal system to maximize social welfare. Part I analyzed determination of the optimal evidence threshold taking other features of the legal system as given, and it also set to the side direct costs of enforcement and of the application of sanctions. Here, we will take resource costs into account and consider interactions among the system's components, viewing the evidence threshold as one of a number of instruments in a system that aims to control harmful behavior. Proof burdens are not often examined from this perspective, but the analysis of Part I makes clear that one of the primary effects of the evidence threshold is on the commission of harmful acts. To be sure, it also influences the chilling of benign acts, which is its other behavioral effect, but so too do other enforcement instruments: Greater enforcement effort, which brings a larger proportion of harmful acts into the legal system, also tends to sweep in additional benign acts. Higher sanctions, unfortunately, apply as well to those mistakenly subject to them. And changes in the accuracy of adjudication generally influence both the likelihood of correctly sanctioning harmful acts and the frequency of mistakenly penalizing benign acts.

Because most of the phenomena examined here have not received significant previous attention and because the number of considerations is large, the analysis will in many respects be preliminary. Furthermore, for ease of exposition, only two moving parts will be considered at a time: in each instance, the evidence threshold and one of the other system components—enforcement effort, sanctions, and accuracy—in Sections III.A, III.B, and III.C, respectively.

138. Moreover, it would be unfortunate if we were indeed reasonably lucky at present, yet the legal system undertook (successful) efforts to push factfinders to adhere more faithfully to the traditional instructions once it was recognized that existing formulations are not closely followed.

A. Enforcement Effort

As introduced in Subsection I.A.1, government enforcement, in the first stage, brings into the legal system a portion of individuals who commit the harmful type of act and also a portion who commit the benign type, all of whom then face adjudication in stage two. It was supposed that acts might be identified by the posting of monitors or auditing, the intensity or rates of which were taken as given—and, for simplicity, the direct costs of which were set to the side. Of course, these enforcement techniques can be applied with different levels of effort, which in turn generate different levels of direct costs: posting more monitors or conducting more audits entails higher expenditures.[139]

One can ask, following a substantial literature on the economics of law enforcement,[140] how the optimal level of enforcement is determined, but now in a context in which there are two types of errors in adjudication. Moreover, it is interesting to inquire into the merits of adjusting enforcement effort versus recalibrating the stringency of the proof requirement in achieving a given level of deterrence. After all, these two instruments are substitutes: for example, one could raise the evidence threshold (which reduces deterrence) and raise enforcement effort (which raises deterrence) in a manner that keeps deterrence constant.[141] Such an exercise will generally change both overall enforcement expenditures and the extent to which benign acts are chilled, potentially raising a different sort of tradeoff from the one examined previously: that between deterrence benefits and chilling costs.

To avoid introducing too many new elements simultaneously, this Section considers only the enforcement costs of, say, posting a number of monitors. (Section IV.A will discuss the additional expenditures required to process more cases in adjudication.) Regarding these enforcement costs, the higher effort required to bring more cases into the system raises legal system expenditures, but there is no offsetting savings from raising the evidence threshold. Put another way, in achieving a target level of deterrence, using additional enforcement effort entails direct resource costs whereas employing a lower

139. Likewise, when enforcement is by investigation—a subject examined in Section IV.B—additional cases can be pursued or they may be investigated more intensively, which entails additional resource costs and results in more individuals being identified in stage one (taking behavior as given).

140. *See, e.g.*, Polinsky & Shavell, *supra* note 102 (surveying the literature).

141. Likewise, if we wished, say, to raise deterrence somewhat, we could inquire into the extent to which this should be accomplished by increasing enforcement effort versus reducing the evidence threshold.

evidence threshold does not. Hence, on this account, a reduced evidence threshold is superior.

If, however, one were to reduce the evidence threshold and enforcement effort so as to keep deterrence constant, chilling costs would not generally remain constant. Instead, chilling costs would tend to rise. Note that, if they did not, the system could not be at an optimum, for reducing the evidence threshold and enforcement effort—keeping deterrence constant—would be unambiguously desirable. Hence, in a well-designed system, it must be that this combination of adjustments increases the chilling of benign acts, creating a tradeoff with the reduction in enforcement expenditures. It remains to explain why this occurs.

To begin, reducing enforcement, taken by itself, will reduce chilling, whereas reducing the evidence threshold raises chilling. Therefore, the suggestion is that the latter effect exceeds the former. To examine this relationship, begin with enforcement effort. In this regard, suppose that, roughly speaking, changing the monitoring intensity changes the rates of identifying harmful and benign acts in stage one by the same proportion.[142] For example, if monitors were increased by 10%, the fraction of both harmful and benign acts brought into the legal system would rise by 10%.[143]

The argument that a lower evidence threshold fares worse than this can be explained both intuitively and graphically. The intuition compares the targeting precision of the two enforcement instruments. As just stated, greater enforcement effort is taken to increase proportionally the flow of cases of each type into the system of adjudication. At a given evidence threshold, the targeting precision on this new flow of cases is the same as that on the

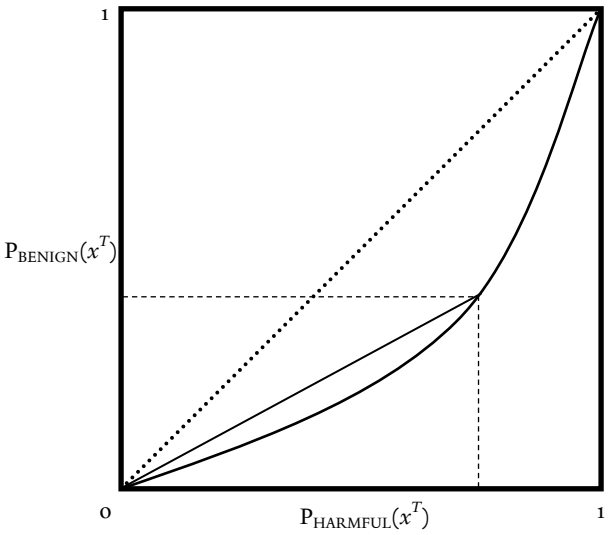
142. A qualitatively different sort of increase in enforcement effort would arise if, say, the number of monitors or audits were held constant but their quality were enhanced (for example, using higher-precision radar in the detection of speeding or employing better-trained auditors). Such expenditures enhance accuracy, the subject of Section III.C.

143. If both fractions were increased by a different amount, say 8%, the analysis would be the same. Note that, as before, nothing is said about the relationship between these two fractions. For example, it might be that, initially, monitors detect 40% of harmful acts and 10% of benign ones; in that case, a 10% rise in the number of monitors would, using the 10% multiplier from the text, increase these rates to 44% and 11% respectively. Of course, these rates need not move in fixed proportions. Specifically, one might suppose that, in prioritizing the placement of monitors, one would first target locations where, ceteris paribus, the relative frequency of harmful acts was highest. In that event, the marginal influence from additional monitors might be relatively less favorable—in targeting harmful rather than benign acts—than the average influence. For random audits, however, these relative proportions might be constant, although audits or inspections in many settings are not entirely random but instead target individuals based on observable indicators of different underlying likelihoods of harmful versus benign acts.

preexisting cases. A lower evidence threshold, by contrast, assigns liability rather than no liability for marginal cases, those where the evidence less strongly indicates that the acts in question were indeed harmful. In particular, the evidence in these marginal cases is weaker than that for those cases previously in the system for which liability was assigned, for all those have levels of x above the preexisting x^T, whereas the newly added liability findings all have an x somewhat below that x^T. Hence, for a common increase in the number of correct findings of liability, an evidence threshold reduction involves a greater increase in the number of mistaken findings of liability than results from raising enforcement effort. This differential is the reason that chilling effects rise when substituting a lower evidence threshold for greater enforcement effort in achieving a given degree of deterrence.

For a graphical exposition that permits a more precise statement, we can use Figure 3 from Subsection I.A.2, which is reproduced here with slight modification as Figure 8.

Figure 8.
$P_{BENIGN}(x^T)$ AS A FUNCTION OF $P_{HARMFUL}(x^T)$ FOR ENFORCEMENT EFFORT–EVIDENCE THRESHOLD COMPARISON



To make the comparison concrete, consider the point on the bowed curve that is now connected to the origin by a solid diagonal line segment. This point denotes the probabilities of applying sanctions (in cases that come to the

tribunal) for harmful and benign acts at some particular evidence threshold x^T.[144] This line segment's slope is given by the ratio of the length of the vertical segment through this point to that of the horizontal segment (each indicated by a dashed line). This slope indicates, for the given evidence threshold x^T, the ratio of the probability that benign acts before the tribunal are subject to sanctions to this probability for harmful acts. Now, as one raises enforcement effort, more cases—some involving benign acts and some harmful acts—are brought before the tribunal and are then subject to sanctions in accord with these probabilities.

The slope of this diagonal may be compared with the slope of the curve at the point under examination. We can immediately see that the curve's slope is greater (the curve is steeper). Now, when the evidence threshold x^T is reduced slightly in order to increase deterrence (moving northeast along the curve), we know that the probability of sanctions for benign acts and the probability for harmful acts both rise. Moreover, starting at the initial evidence threshold x^T, they rise in the ratio given by the slope of this curve. After all, the curve in the figure was constructed to depict precisely this relationship, as explained previously. It follows, therefore, that when deterrence is enhanced somewhat by reducing the evidence threshold x^T from some initial point, rather than by raising enforcement effort, the probability of applying sanctions to benign acts will rise relatively more. The reason the curve plausibly has this shape is rather subtle: as explained at the end of Subsection I.A.2, it relates to the underlying meaning of stronger versus weaker evidence. And this meaning, in turn, corresponds to the intuition given before presenting Figure 8.

To summarize, it turns out that the choice between the use of greater enforcement effort and of a lower evidence threshold to achieve deterrence involves a tradeoff—between expending more resources and imposing larger chilling costs—all in achieving some given or target level of deterrence. Optimal system design will, accordingly, balance these considerations. If more enforcement is relatively cheap, it will be better to use high enforcement effort and a strict evidence threshold. But if greater enforcement is quite costly, the opposite mix makes sense. Likewise, the greater the magnitude of incremental chilling costs, the more it is appropriate to achieve deterrence through tougher

144. Inspection of Figure 8 reveals that, for any point we might have picked on the bowed curve except the corner points, the relationship between the slopes would be as presented in the text to follow. (The lower left corner, where all individuals are exonerated no matter how strong is the evidence, corresponds to an infinite evidence threshold, and the upper right corner, where all individuals are sanctioned no matter how weak is the evidence, corresponds to a never-binding or zero threshold. Actually, the property relating the slopes also holds at the upper right corner, although there it is not possible to further reduce the evidence threshold in a meaningful way.)

enforcement rather than through a lax evidence threshold, and conversely when chilling costs are mild.[145]

This set of prescriptions, on reflection, accords with intuition—if the objective in designing the legal system as a whole is to maximize social welfare—even though this seemingly obvious and natural tradeoff depends on factors that are more elusive than one might have suspected. The main purpose of the analysis in this Section is to highlight a design choice that is not often considered explicitly. Part I analyzes determination of the optimal evidence threshold when other features of the legal system, like enforcement effort, are taken as given. But the evidence threshold is a component of the broader system, so when considering how in principle it should be set, it makes sense to also have on the table the question of how other enforcement instruments should be calibrated.

B. Sanctions

1. Level

Before turning to the interaction between setting an optimal evidence threshold and determining the appropriate level of sanctions, it is useful to begin with some remarks on the optimal determination of enforcement effort and sanctions, which has been a significant topic in the economics of law enforcement literature, although usually in settings in which errors are not part of the analysis. Specifically, as Gary Becker argues, for any given, nonmaximal level of sanctions, it tends to be optimal to raise sanctions and reduce enforcement effort.[146] If this is done so as to keep deterrence constant, one saves enforcement resources while, in simple settings, having no other effects. The literature then explores various reasons why this prescription may not hold.[147]

145. Just as when determining the optimal evidence threshold taking enforcement effort as given, it is clear that the socially best mix of enforcement effort and evidentiary requirements will depend greatly on the context. Indeed, even if the optimal evidence threshold for a given, common level of enforcement would have been the same in two settings, it is quite possible that the optimal thresholds will differ substantially when enforcement costs are significantly different and it is possible to alter both enforcement effort and the evidence threshold, tailoring each to the particular circumstances.

146. *See* Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169, 183-84 (1968).

147. *See, e.g.*, Polinsky & Shavell, *supra* note 102, at 414 n.19, 415-16, 431-34.

In this regard, it is natural to inquire whether the presence of errors upsets this basic argument. As a first cut—and taking the evidence threshold as fixed—the answer is negative.[148] Regarding mistaken exoneration of individuals who in fact committed harmful acts, it is true that, as a consequence of errors, a lower portion are sanctioned. But if, say, enforcement effort is reduced so as to cut in half the flow of individuals into the legal system, and the sanction is simultaneously doubled, the expected sanction on individuals contemplating harmful acts remains the same. The presence of errors—say, adjudication mistakenly exonerates 20% of individuals who actually committed harmful acts—means that the ex ante probability of sanctions is 20% lower in both instances, but a 20% reduction in two numbers keeps their relative proportion unchanged.

The analysis is the same regarding the mistaken imposition of sanctions on individuals who committed benign acts. Supposing for concreteness that this error rate is 10%, we would, as in the preceding illustration, have a 10% reduction in the ex ante probability of sanctions (now, for benign acts) in both the higher and lower enforcement effort scenarios. Given the hypothesized reduction in enforcement effort, which halves the number of individuals who are subject to adjudication, the ex ante likelihood that individuals who contemplate benign acts would thereby be sanctioned is cut in half. And, when they are sanctioned, they face double the sanction. As a consequence, their expected sanction is likewise the same. More broadly, it is not immediately obvious that making bigger mistakes less often is worse than making smaller mistakes more often; and in the basic setting under examination, these proportions are the same, so the costs of mistakes in terms of reductions in social welfare are the same. In sum, both deterrence and chilling effects are unaffected by this policy experiment of halving enforcement intensity and doubling sanctions, but enforcement resources are saved. (This core argument may or may not be affected when additional complications are introduced.[149])

148. For prior analysis of this point, which implicitly takes the burden of proof to be fixed, see Kaplow, *supra* note 102, at 349-50, and Kaplow & Shavell, *supra* note 102, at 5-6.

149. Many can be imagined. Consider some involving costly sanctions, the subject of Subsection III.B.2. One possibility is that sanctions involve social costs of imposition that are proportional to the sanctions imposed. In that case, there would be no direct effect of sanction costs: after all, although sanctions are twice as high, they are imposed half as often; hence, the total imposition of sanctions is constant. Another possibility is that sanction costs are nonlinear, as arises with risk aversion in the case of monetary penalties. *See, e.g.*, Michael K. Block & Joseph Gregory Sidak, *The Cost of Antitrust Deterrence: Why Not Hang a Price Fixer Now and Then?*, 68 GEO. L.J. 1131, 1135-38 (1980) (discussing risk aversion and error); A. Mitchell Polinsky & Steven Shavell, *The Optimal Tradeoff Between the Probability and Magnitude of Fines*, 69 AM. ECON. REV. 880 (1979). The most direct implication is that, to

Consider next our prime concern, the interaction between setting the evidence threshold and the level of sanctions. Here, the analysis has some features in common with Section III.A's analysis of enforcement effort, but there is also a notable difference. Recall that raising enforcement effort and raising the evidence threshold simultaneously, in a manner that holds deterrence constant, entails additional expenditures but reduces chilling costs. If one instead raises the level of sanctions and the evidence threshold, holding deterrence constant, there is no longer any increase in enforcement expenditures but there still is a reduction in chilling costs. To understand this result more fully, note first that raising sanctions and raising enforcement effort are taken to hit both groups—individuals who committed harmful acts and those who engaged in benign ones—in the same proportion. As a consequence, this experiment of raising sanctions and the evidence threshold in a manner that keeps deterrence constant manages to hold expenditures on stage-one enforcement constant (because enforcement is unchanged, unlike before) but also to reduce expected sanctions on those who commit benign acts (because in this respect the present experiment acts just like that with an increase in enforcement), making the overall welfare impact positive.[150]

The result that consideration of errors—and in particular, errors involving the mistaken imposition of sanctions on individuals who committed benign acts—generates an argument in favor of high sanctions may seem surprising.

maintain a given level of deterrence when the probability of sanctions is cut in half, the sanction need not double but can be raised by a lesser amount. Risk-bearing costs are greater, which may make the policy revision welfare reducing overall. Note that, in any case, there is a parallel between the increase in risk-bearing costs imposed on both sets of individuals, in that both face the same higher sanction. The amount by which the sanction must increase does not depend on the presence of error per se, but rather (since deterrence is being held constant) only on individuals' degree of risk aversion, which dictates how much monetary sanctions must increase to double the disutility of sanctions. The main difference is that, with errors, sanctions are imposed more frequently at any given level of deterrence because they also fall on some individuals who commit benign acts, so sanction costs become relatively more important. For prior analyses of how sanction costs influence optimal enforcement effort and sanction levels—in settings that do not examine errors—see, for example, Louis Kaplow, *A Note on the Optimal Use of Nonmonetary Sanctions*, 42 J. PUB. ECON. 245 (1990); Louis Kaplow, *The Optimal Probability and Magnitude of Fines for Acts that Definitely Are Undesirable*, 12 INT'L REV. L. & ECON. 3 (1992); Polinsky & Shavell, *supra*; A. Mitchell Polinsky & Steven Shavell, *The Optimal Use of Fines and Imprisonment*, 24 J. PUB. ECON. 89 (1984).

150. Taking the evidence threshold—in addition to enforcement effort—to be an adjustable enforcement instrument thus offers an additional type of argument paralleling Becker's in favor of higher sanctions. Moreover, because it operates differently, it can apply in contexts in which Becker's original argument does not hold, and vice versa. *See* Kaplow, Optimal Burden of Proof, *supra* note 16.

But, first, we already saw that, in a simple and direct sense, such mistakes do not in fact favor lower sanctions because, if one wishes to maintain deterrence, the shortfall must be made up with greater enforcement effort, which increases the overall rate of mistakes. Second, an advantage of higher sanctions is that they allow us to reduce the legal system's pressure on harmful acts in other ways, one of which is by enabling a stricter evidence threshold. And this tougher proof requirement, in turn, means lower expected sanctions for benign acts, even taking into account the contrary but weaker effect thereon due to the higher sanctions.[151]

This observation illustrates the broader theme of this Part: optimal legal system design must not only consider each component in isolation but also needs to take into account interactions among them. Before concluding the present discussion of the level of sanctions, note that this Subsection does not purport to suggest that optimal sanctions are maximal; it only shows that the influence of error and consideration of how the evidence threshold might be set do not provide a simple, strong basis for employing low sanctions. Explanations will need to be found elsewhere, either in more subtle and complex considerations involving error and proof burdens or in other factors entirely.

2. Nonmonetary Sanctions

Until now, sanctions themselves have been taken to be socially costless, as is the case when monetary sanctions (fines or damages) are imposed on risk-neutral individuals—where defendants' payments are offset by receipts by the state or private plaintiffs.[152] Consider instead nonmonetary sanctions, such as imprisonment, that are costly in themselves: neither the loss of liberty nor the

151. Another consideration is suggested by Kaplow, *supra* note 102, at 351-52, and Kaplow & Shavell, *supra* note 102, at 5-6, who observe that if one raises sanctions and reduces enforcement effort so as to keep deterrence constant, there are fewer cases in adjudication and hence greater accuracy (the subject of Section III.C) can be achieved with lower expenditures, providing another channel by which error costs may be reduced.

152. Some discussion of sanction costs was offered in addressing errors and the level of sanctions in the preceding Subsection, in note 149. Note that even with monetary sanctions and risk neutrality, there is a further implicit assumption that collecting sanctions is costless. If instead such costs were, say, proportional to the sanction, then the analysis of this Subsection would apply directly. If collection costs were fixed, independent of the magnitude of the sanction, the analysis would differ, having some features in common with audit costs, examined in Section IV.A. The difference is that audit costs are, in the simplest case, proportional to the number of acts whereas this sort of sanction cost would weight the relevant numbers by the rate at which sanctions are ultimately imposed.

costs of operating prisons are mere transfers. When sanctions are socially costly, it tends to be optimal to reduce their imposition, which may seem to favor a higher evidence threshold.[153] While the initial hypothesis is correct, this conclusion need not follow because the evidence threshold affects behavior.[154]

To begin the analysis, if behavior is taken as given, a higher evidence threshold reduces both the rate of (correct) imposition of sanctions on individuals who committed harmful acts and the rate of (incorrect) imposition of sanctions on individuals who committed benign acts. Moreover, both reductions are, in themselves, advantageous on account of the reduction in social sanction costs. But this consideration is only part of the story.

As is familiar by now, a higher evidence threshold also affects behavior. Specifically, by reducing the expected frequency with which sanctions are imposed on both types of actors, a greater number of individuals will commit harmful acts and benign acts. Both effects increase the number of individuals who enter adjudication, and hence both increase the frequency of imposition of sanctions and, thereby, the magnitude of social sanction costs.

A priori, either effect could be greater: the increased flow of cases could raise the total imposition of sanctions by less than the decreased likelihood of sanctions in a given case reduces the total, or vice versa. It seems plausible to suppose that in some situations the former will be dominant and in others the latter. Hence, introducing social costs of sanctions, such as those that arise with nonmonetary sanctions, will sometimes favor a higher evidence threshold and in other instances a lower one. Likewise, factors indicating that social sanction costs are greater than might previously have been contemplated will, in general, have an ambiguous effect on the optimal height of the evidence threshold.

Note, moreover, that similar logic applies with regard to possible asymmetries in sanction costs. For example, suppose that the direct social cost of imposing each unit of a sanction—behavioral effects aside—is greater with regard to individuals who committed benign acts and bear the sanction by mistake than with regard to those who committed harmful acts and properly bear the sanction.[155] Again, we have a tradeoff: a higher evidence threshold

153. *See, e.g.*, POSNER, *supra* note 7, at 828 (arguing that the cost of imprisonment favors erring on the side of acquittals, thereby providing a rationale for the criminal law's requirement of proof beyond a reasonable doubt).

154. For analysis of the issue in a simpler model without chilling effects, in which the analysis is somewhat different, see Kaplow, *supra* note 102, at 359-62.

155. It is not obvious that this is so, and one can imagine various reasons that cut in different directions. To be sure, the psychological cost of imprisonment may be greater for those who are there by mistake. On the other hand, the criminogenic effect of imprisonment may be

reduces the likelihood of imposition of the more costly sanction on any given individual who committed a benign act and is before the tribunal, but the resulting reduction in the expected sanction for benign acts implies that fewer of them are chilled, so a greater number of such individuals come before tribunals and thus may bear sanctions by mistake. Either effect could be larger. In other words, it does not follow that a higher social cost associated per se with the mistaken imposition of sanctions on innocent individuals favors a higher evidence threshold. The opposite is also plausible, and which situation prevails in any given setting is an empirical question.

It is worth reflecting briefly on why many ordinarily imagine that a heightened evidence threshold is favored by a greater social cost being associated with the mistaken imposition of sanctions on individuals who have not committed harmful acts. It seems likely that this supposition is generated by an ex post view of the legal system, which takes as given the cases that come before the tribunal. This perspective is closely related to those underlying conventional conceptions of the burden of proof that were the subject of Part II. In considering the question, onlookers presumably contemplate an actual case, one in which there is uncertainty about the truth with regard to the individual before the tribunal. In considering what decision is best, a greater cost associated with the mistaken imposition of sanctions naturally favors a tougher proof burden. We can now see that this perspective overlooks the ex ante effect on behavior that works in the opposite direction and can result in a net increase in the mistaken imposition of sanctions.[156]

larger for those who did commit harmful acts. Yet another factor is that there may be incapacitation benefits regarding those who commit harmful acts but no such effects (or reduced ones) for those who did not. Such benefits reduce the net social cost of imprisonment and, if sufficiently strong, can make the net cost negative, that is, a net benefit, with regard to individuals who committed harmful acts.

156. *See supra* notes 103, 112. Interestingly, there is also a tension within this behavioral channel: ordinarily, we suppose that the chilling of benign acts is undesirable, but here chilling also reduces the imposition of socially costly sanctions. Now, from the ex ante perspective of individuals who contemplate committing benign acts, they are better off if the expected sanction (in terms of the expected overall utility cost) is reduced. Moreover, if this leads them to commit a benign act that otherwise would have been chilled, it is necessarily the case that, prospectively, they are better off. It is possible, however, that the net effect, focusing solely on this shift, is socially undesirable as a consequence of externalities, the most obvious of which in the present context is that individuals do not directly bear the costs of operating prisons. (Society as a whole does, but a particular individual contemplating an act only bears, through taxes, a negligible fraction of the expected increase in the cost of operating prisons that flows from committing the act in question.)

C. Accuracy

Throughout this investigation, the quality of the evidence is taken as given, so the only system design question regarding evidence is how strong it must be in order to assign liability. However, myriad features of the legal system influence its accuracy, regarding both the actual information that is generated and also the care with which it is assessed (another feature of quality, really).[157] Many rules of legal procedure, including laws regulating evidence and governing appeals, affect accuracy, as do policies affecting the ability, available time, and incentives of investigators and lawyers.[158] Because the accuracy of the legal system determines the error rates associated with a given proof requirement, and these error rates are key determinants of the optimal evidence threshold,[159] it is appropriate to inquire into the relationship between accuracy and the burden of proof.[160]

As a preliminary matter, it is important to distinguish the two concepts, which are sometimes confused because many changes in system design simultaneously influence both, even if not formally.[161] Consider the effects of providing additional investigative resources to indigent criminal defendants. Suppose that the result is to improve the overall quality of information available to the tribunal; better information may be associated with greater accuracy. However, in this instance there may also be a de facto strengthening of the burden of proof: even if the requisite minimum strength of evidence for conviction is literally unchanged,[162] investigators working for defendants will

157. By analogy, consider the use of imaging for medical diagnosis. The quality of information can be improved through a higher-resolution scan or by employing a more highly skilled interpreter (which itself may be achieved, de facto, in many ways, including better training, more time, and the use of second opinions).

158. For example, when Section III.A explored raising enforcement effort, it was supposed that this was accomplished by posting more monitors or auditing more actors. But one could instead consider improving the quality of monitoring, auditing, inspection, investigation, and so forth, in which case the analysis of the present Section would be applicable.

159. More precisely, what matters is how these error rates change with the evidence threshold.

160. This inquiry also prompts a comparison of prior literature on the subjects by commentators. When researching an extensive treatment on accuracy in adjudication, Kaplow, *supra* note 102, I similarly discovered that most legal analysis did not really attempt to examine the consequences of different levels of accuracy for how well the legal system functions. That is, there was also essentially a vacuum with regard to the rationales that underlie the relevant tradeoff, in that context between accuracy and system cost.

161. *See id*. at 357-59.

162. There are different ways to understand whether the implicit change in the proof burden should be understood as merely de facto or also de jure. For a given evidence threshold x^T,

be selective both in where they focus their efforts and in what subset of their findings are presented to the tribunal.[163] As a consequence, it will on average be more difficult to establish guilt—for individuals who committed benign acts and also for those who committed harmful acts.[164]

This entanglement of accuracy and the proof burden can helpfully be separated. Specifically, one can combine a proposed intervention that affects both—such as in our example—with a simultaneous adjustment to the formal evidence threshold that restores the prior de facto proof burden. This clarifies the difference between the two phenomena, which is useful conceptually. It also has pragmatic appeal. For example, if the intervention is desirable on accuracy grounds (perhaps important information that may exonerate innocent defendants is often overlooked) but not on proof burden grounds (perhaps the previous balance was just right), then it would be sensible to implement the change but adjust the evidence threshold so as to hold the de facto proof burden constant. On the other hand, if little improvement in information results, so that the reform is not cost-justified on accuracy grounds, but it was motivated by a desire to raise an excessively lenient proof burden, it would be preferable to accomplish the latter directly and forgo the initial reform.

Having distinguished the concepts of accuracy and proof burden, consider next how the degree of accuracy affects the optimal evidence threshold.

---

one might view a general enhancement of presentations by the defense as reducing the distribution of evidence strengths x for both types of actors (shifting both curves in Figure 1 to the left), in which case the de facto proof burden would be higher (met less often) even though the de jure burden is the same. Alternatively, one could say that the de jure burden has changed because the meaning of evidence strength and thus of a given x^T has shifted. Nothing in the present analysis depends on this difference in interpretations.

163. On the selection of evidence in adjudication, see, for example, Louis Kaplow & Steven Shavell, *Legal Advice About Information To Present in Litigation: Its Effects and Social Desirability*, 102 HARV. L. REV. 565 (1989).

164. Note that if, instead of stating the evidentiary requirement in terms of an evidence threshold x^T, one employs a conventional burden of proof notion of the sort examined in Part II—a formulation based on ex post likelihoods—then a Bayesian factfinder would implicitly adjust the evidence threshold x^T in order to keep constant the requisite likelihood. In that sense, there would be no implicit change in the proof burden. To see this, the additional investigative resources might be seen as shifting both curves in Figure 1 to the left. For any level of x, there will accordingly be associated a different probability than before. Hence, when examining the lower elements of the boxes in Figure 5, it would generally be true that different values would be assigned for any contemplated x^T. For concreteness, consider the case in which the shapes of the two curves in Figure 1 remain the same and they shift left by precisely the same amount. Then, whatever x^T previously worked in Figure 5, it will now be true that a new x^T that is lower by precisely the amount of that leftward shift will work (because it will be associated with the same level for both probabilities as prevailed—and worked—initially).

Suppose, for example, that the system is redesigned in some fashion so as to increase accuracy at the existing evidence threshold, which is to say that there is a reduction in both error rates: the likelihood of mistakenly exonerating those before tribunals who committed harmful acts and the likelihood of mistakenly sanctioning those who committed benign acts.[165] Is the optimal evidence threshold in this new regime higher or lower?

It turns out that there is no simple or unambiguous answer to this question. First, it is difficult to give it sharp meaning: that is, it is hard to compare the stringency of evidence thresholds across regimes with different levels of accuracy. To illustrate some of the problems, suppose that in the more accurate regime we begin with an evidence threshold that involves the same rate of mistakenly sanctioning individuals before the tribunal who committed benign acts—which, since accuracy is higher, entails that the rate of mistaken exoneration is lower. In that case, it may be optimal to raise the evidence threshold somewhat. But if we begin in the new regime at a threshold that involves the same rate of mistaken exoneration as before—which entails a lower rate of mistaken sanctioning—it may be optimal to lower the threshold somewhat. Supposing that both are true, it may seem natural to consider points in between, but depending on the starting point chosen, it may be optimal to raise the evidence threshold, lower it, or keep it as is. To draw an analogy, suppose that for medical diagnosis we switch to a new, more accurate test meter, whose accuracy is generated by utilizing a different underlying type of test. There simply may be no natural, inherent relationship between the scales of the new meter and the old one, so there is no clear basis for saying what cutoff reading should be viewed as corresponding to what was used before.[166]

Second, putting this definitional question to the side, as a matter of substance there are multiple and conflicting effects. Consider a starting point

165. One way to define greater accuracy is to consider a regime such that the curve in Figure 3 is everywhere (except at the corner points) closer to the lower right corner of the box. This depiction immediately suggests a further complication: one can imagine reforms that would raise accuracy for some thresholds (perhaps low ones, corresponding to the upper right portion of the figure) but reduce it for others (perhaps high ones, corresponding to the lower left portion of the figure). In that case, the new curve would be lower in the upper right portion and higher (closer to the 45° line) in the lower left.

166. One could align proof burdens by initially imposing in the new regime the same ex post likelihood that the individual before the tribunal committed a harmful act as prevailed in the old, and then asking how it should optimally be adjusted. However, that calibration is just one possible choice, and the analysis in Subsection II.C.1 suggests that it could have strange properties because it depends on behavior, which in turn is influenced by the degree of accuracy and the evidence threshold that is chosen.

in the new regime such that the rate of both types of errors is lower. A reduced rate of mistaken exonerations means that the expected sanction facing those contemplating harmful acts is higher, which improves deterrence. As explained in Section I.B, when deterrence is higher, the marginal benefit of further deterrence is lower, so on this account the optimal evidence threshold would rise. On the other hand, a lower rate of mistaken sanctions reduces the expected sanction facing those contemplating benign acts, which reduces chilling. Section I.B further explained that, when chilling is lower, the marginal cost of further chilling is reduced, so on this account the optimal evidence threshold would fall.

Examination of Figure 4 or 6, indicating the determinants of the optimal evidence threshold, reveals other possible differences, the preceding argument having addressed only the bottom elements in the two boxes. The middle elements, the concentrations of marginal harmful and benign acts, could each be higher or lower, depending on the distributions of individuals' private benefits for the two types of acts. And even the top elements could differ, depending on what starting point for the evidence threshold is chosen in the new regime and on the ratio between the rates of change in the probabilities of sanctions for harmful and benign acts under the new regime versus the old.[167]

In sum, raising accuracy has many effects on the determinants of the optimal evidence threshold; some point in opposite directions while others have ambiguous implications. Accordingly, little can be said about how changing the degree of accuracy affects the optimal threshold. A corollary is that it is difficult to say how evidence thresholds should vary across legal systems with differing degrees of accuracy.

A final question of interest is how the analysis in the present investigation illuminates the important but rarely addressed question of how one can determine the social value of accuracy in a legal system—that is, how many additional resources it makes sense to expend in order to achieve some degree of improvement in accuracy.[168] Regarding this inquiry, a fairly direct method of valuation is provided. Consult again Figure 4 or 6: the left box indicates the deterrence benefit of a reduction in the evidence threshold and the right box

167. As explained in note 165, greater accuracy can be depicted as the curve in Figure 3 being bowed further toward the lower right. Depending on where one starts on the new curve, the curve's slope—which indicates the ratio described in the text—could be higher than, lower than, or the same as the slope on the original curve at the original evidence threshold.

168. Kaplow, *supra* note 102, and Kaplow & Shavell, *supra* note 102, address accuracy in the determination of liability, but in a setting that does not include chilling effects, so the analysis is quite different.

the chilling cost. With some reinterpretation, these same boxes indicate the effects of an increase in accuracy.

To see this point, suppose that we reduce the two types of error by some amounts but hold everything else in the system constant. There will be an increase in deterrence benefits and a reduction in chilling costs (not an increase, as when relaxing the evidence threshold). To quantify the former, we can start with the top element in the left box. Using our analysis from Section I.B, it is straightforward to determine the increase in the expected sanction for harmful acts: we again have an increase in the probability of being subject to sanctions when before the tribunal, and all else is constant, so the expected sanction for harmful acts rises in proportion to the rise in this probability on account of the increase in accuracy. The other elements in the left box are the same, so the deterrence benefit is calculated as before.

For the chilling benefit, we look to the right box and repeat the exercise. The greater accuracy implies a fall in the expected sanction for benign acts, proportional to the drop in the probability of benign acts being sanctioned when before the tribunal. The other elements in the right box are the same, so we can calculate the chilling benefit.

Taken together, we have a deterrence benefit and a chilling benefit. The sum of these gains is the value of the increase in accuracy, and this total can be compared to the resource cost of the improvement to accuracy to determine whether that expenditure is desirable—that is, whether it raises social welfare. Although, as with the determination of the optimal evidence threshold itself, these calculations will be difficult because the information on the determinants is hard to come by, the correct conceptual formulation is evident.[169]

IV. ALTERNATIVE METHODS OF ENFORCEMENT

Until now, little attention has been given to the particular method of enforcement or to how different methods may have different implications regarding how the evidence threshold should be set. For ease of exposition, it has been assumed that enforcement sweeps in certain fractions of acts of each type and that greater effort raises those proportions. This analysis most closely corresponds to what Section IV.A calls monitoring. This context was chosen because it is less complex than some of the others and because its core features are present under other techniques when ex ante behavior is central. This Part

169. Moreover, as discussed in Subsection II.B.2, challenges in applying the correct mode of assessment do not justify substituting alternative, easier criteria unless they are reasonably good proxies for what really matters.

considers alternatives. Section IV.A discusses auditing alongside monitoring, showing their similarities with regard to effects on behavior—deterrence and chilling effects—but differences with regard to enforcement costs. Section IV.B examines investigation, which is qualitatively distinct on both dimensions. Section IV.C explores a highly dissimilar enforcement setting, the regulation of future conduct (for example, licensing), wherein legal rulings dictate what parties are permitted to do going forward rather than being chiefly concerned with affixing sanctions in light of past behavior. Finally, Section IV.D considers how the evidence threshold may influence prosecutors' and plaintiffs' behavior and, as a consequence, what cases come before the tribunal in the first instance, even taking individuals' acts as given.

It is worth observing at the outset that varying enforcement techniques are sometimes used interchangeably or in a complementary fashion even when addressing a single problem. For example, with some environmental regulation, there may be audits through spot checks and investigations when contamination is detected, as well as regulations that dictate specific future conduct; likewise, harmful activities may be subject to both public and private enforcement. By contrast, in other settings, a single mode may be primary. For example, with arson, enforcement is mainly through investigation that is undertaken by public authorities.

A. Monitoring and Auditing

This Section begins with monitoring, the focus until now, and then analyzes auditing. As mentioned, the two modes of enforcement are similar with regard to behavior but differ in terms of enforcement costs.

Monitoring may be implemented, for example, by posting police officers along a highway for the purpose of detecting speeding or through the use of patrols on the lookout for a variety of violations. There are private analogues as well. Stores may employ security guards to detect shoplifting, large property owners may post monitors for trespassing, and intellectual property owners may task some employees to keep a lookout for infringement. Monitors are presumed here to detect some fraction of individuals who commit harmful acts and also a fraction (perhaps much lower) of individuals who actually commit benign acts, and these are the acts that enter adjudication. As a simplification and rough approximation for some settings, it is imagined that, although posting more monitors to raise the detection rate (for harmful acts and, as a byproduct, for benign acts) is more expensive, as analyzed in Section III.A, the cost depends on the number of monitors and not on how many apparent violations they observe. For example, a security guard will spend most of the time just watching, and only occasionally will detect what seems to be a

perpetrator, but the cost of employing the guard will be essentially independent of whether one shoplifter or five are identified per week.

Compare auditing. This technique is perhaps most familiar from its use in enforcing income tax laws, but it is also deployed much more widely. Random inspections are used to check for violations of health laws, fire regulations, environmental edicts, and occupational safety measures. Private parties engage in similar practices, such as when they examine some inputs from suppliers to check for defects. The essential character is that a given proportion of all activity is examined, and it costs more (perhaps proportionally more) to examine a higher proportion. When an audit or spot check is conducted, some fraction each of harmful and benign acts (presumably, a greater percentage of the former) will be detected as involving apparent violations, and these acts enter adjudication.

Regarding effects on behavior, monitoring and auditing are quite similar. A given intensity of enforcement is associated with the identification of some portion of each of the two types of acts, and raising enforcement effort increases these portions. Of course, the particular fractions will depend on the setting, and in some cases we might expect one or the other technique to have greater targeting precision or to be more expensive for a given hit rate. When it comes to setting the evidence threshold, therefore, the numbers for each element in the boxes in Figures 4 or 6 may well differ, so different thresholds may be optimal with regard to behavioral effects on social welfare. But, qualitatively, the analysis is the same.

However, behavioral effects on welfare are not all that is relevant; enforcement costs must also be taken into account. In Part I, these costs were ignored. For monitoring, that omission was inconsequential because enforcement effort and thus the cost of monitoring was constant—in particular, it was unaffected by the evidence threshold. However, even as an approximation, this constancy of enforcement costs does not extend to auditing.

Here's the difference: As noted, with monitoring, the cost is taken to be the same whether one or five apparent violations is observed by the posted agent. But, with auditing, even holding the audit rate constant, say, at ten percent, costs depend on individuals' levels of activity. For example, if more restaurants operate (whether properly or not), more health inspections need to be conducted—although the percentage audited remains the same. If more manufacturing is undertaken, there need to be more safety inspections. And so forth.

The implication is that, for auditing, the evidence threshold directly affects audit costs. Specifically, when the evidence threshold is reduced somewhat—our thought experiment from Part I—we know that there is more deterrence

and more chilling. Both involve reductions in activity: harmful and benign acts fall. Because there are fewer acts, fewer audits need to be conducted at our fixed audit rate, so audit costs are lower. Hence, a lower evidence threshold is more desirable than one would have concluded considering solely the direct social welfare consequences associated with behavior.

Reflecting on the foregoing, we can see that deterrence is more valuable with auditing than with monitoring, because in addition to avoiding the external harm caused by harmful acts we also reduce expected audit costs. Moreover, chilling is less undesirable; that is, what was previously a pure cost—losing the benefits of benign acts—is now less of a social cost because of the partial offset from saving audit expenses. It is even possible that chilling could be beneficial. This seemingly unexpected conclusion is not, however, very surprising on reflection. Consider, for example, an activity that can be highly dangerous when not conducted properly, which as a consequence makes it optimal to employ a high audit rate; in some settings, the rate is essentially one hundred percent for this reason. It would be socially advantageous for such activities not to be conducted if they are at a very small scale, even if the operator would in fact behave in a benign manner, for the expense of performing the necessary audits could exceed any benefit from the activity. In fact, for some dangerous activities, fees must be paid to cover inspection costs, and presumably such fees discourage operations on so small a scale that the fees cannot be recovered.

Paralleling the discussion in Section II.B, we can compare the determinants of the optimal evidence threshold when enforcement is by auditing with the determinants of the preponderance of the evidence rule and other rules based on the ex post likelihood that the individual before the tribunal committed a harmful act. Again, the contrast is striking. Regarding the direct effects of behavior on welfare, we have all the differences we had before. And now we have additional divergences because the enforcement cost effects just described are also omitted from the preponderance rule and others like it,[170] reinforcing the point that conventional conceptions of the burden of proof are not even proxies for the social welfare effects of changing the evidence threshold.

170. It is true that the frequencies of the two types of acts and the audit rate are relevant under the preponderance rule (they are components of the upper elements in the boxes in Figure 5) and in the calculation of total audit costs. However, these factors are used in an entirely different manner. For the preponderance rule, the two frequencies (multiplied by the common audit rate and also multiplied by different hit rates) are on opposite sides of an equation, whereas for audit costs these frequencies (multiplied by the audit rate and also by the cost per audit) are added together.

Finally, note that this discussion of audit costs and the preceding discussion of monitoring costs oversimplify by omitting processing costs that follow initial targeting. In both instances, once individuals are identified—correctly or not—in stage one, they enter into the legal system, ultimately resulting in adjudication, and this second stage is itself costly.[171] Therefore, both modes of enforcement entail yet another cost, one that depends on enforcement intensity, the number of the two types of acts that are committed, and the hit rates for each of the two types of acts. Because one component is, again, the number of acts committed, there is a further benefit to deterrence and to chilling (perhaps a partial offset to the chilling cost) as a consequence of reducing the evidence threshold.[172]

B. Investigation

Enforcement by investigation differs substantially from enforcement by monitoring or auditing and in ways that have important implications for the determination of the optimal evidence threshold. Investigation refers here to enforcement triggered by the observation of a harmful act. This event instigates follow-up activity designed to ascertain responsibility. Such an investigation with some probability identifies the individual who committed the harmful act, and with another probability (hopefully, a much lower one) it identifies an individual who actually committed a benign act. With most instances of murder or automobile theft, for example, it may be readily apparent that a harmful act has occurred but require investigative resources to identify the perpetrator. There are civil analogues, most obviously regarding many torts, where again the fact of injury may be clear.

When enforcement is by investigation, there is an important difference in the manner in which harmful acts versus benign acts may give rise to an individual actor ending up in adjudication. One contemplating a harmful act anticipates that the act will result in an investigation (with near certainty or probabilistically, depending on the setting), which might in turn identify the actor and lead to adjudication. By contrast, in many settings the commission of

171. If one considers that many disputes settle (plea bargaining in the criminal context), these costs are lower than otherwise, but they are still positive and, in many settings, significant.

172. This factor is yet another difference between the determinants of the optimal evidence threshold—that which maximizes social welfare—and of the preponderance rule and others like it. As explained in note 170, we again have some overlap in components, but for assessing welfare the quantities associated with harmful acts and benign ones are added rather than balanced against each other, as they are under the preponderance rule.

a benign act will not trigger an investigation.[173] So why, then, would individuals who commit such benign acts nevertheless face adjudication in some instances? This possibility arises in the present context because of actions that may lead an individual to be confused with an actual perpetrator. Driving a van at night in the vicinity of a high-burglary neighborhood might lead one to be mistakenly apprehended for a break-in. Purchasing a significant quantity of used electronics from an individual may lead to suspicion of complicity if that individual is part of a band of thieves. Being a relative or acquaintance of someone with enemies, especially if one is known to be among those enemies, may lead one to be accused if the person is harmed by someone else. In the civil context, the manufacture of products that are quite safe but are used in conjunction with others that are not could result in products liability allegations. These are the sorts of actions that might be chilled by the prospect of mistaken identification, followed by a finding of liability.

This contrast between the present setting and that considered with monitoring or auditing has an immediate and important implication: anything that contributes to the deterrence of harmful acts will, as a direct consequence, reduce the expected sanction for benign acts and thus chilling costs. The reason is that harmful acts are what trigger the investigations that sometimes misfire and lead to sanctions being imposed on individuals who commit benign acts. When there are fewer harmful acts, fewer investigations are triggered,[174] and—for a given misfire rate—fewer instances of mistaken imposition of sanctions will result. This is not the only impact of policies that raise deterrence, but it is one effect. Note also that the reduction in investigations will save enforcement resources, another benefit of deterrence.

Consider the determinants of the optimal evidence threshold when enforcement has these features. Figures 4 and 6 depict the components of

173. Often in civil settings and sometimes with crime, uncertainty may exist not with regard to who caused harm but rather about whether there was improper behavior. For example, it may be clear to a buyer that goods were not delivered but require further investigation to determine whether the seller was at fault and thus in breach of contract. And with some crimes, an actor may be excused from causing injury, for example, on account of self-defense. As will become clear, the analysis of these situations would not differ as significantly from the cases with monitoring or auditing because one who causes harm even when acting properly may thereby trigger scrutiny.

174. If, say, a police force has a fixed budget for investigations, this reduction in the number of investigations might not occur. However, over time, if there are fewer crimes to investigate, one would expect budgets to fall. Moreover, even in the short run, if there are fewer crimes, one response is likely to involve more intensive investigation of remaining cases (rather than a greater number of investigations), and the enhanced depth may on average improve the accuracy of the results, a subject that was explored in Section III.C.

deterrence benefits and chilling costs when enforcement is by monitoring or auditing. When enforcement is by investigation, there are a number of differences. Starting with deterrence benefits, the lower element in the left box in the figures would have two supplements, corresponding to the two additional effects just explained: each act deterred, in addition to avoiding the external social harm of the act itself and forgoing the private benefit from the act, now also reduces chilling costs and enforcement costs, both as a consequence of there being fewer investigations. These factors favor a lower evidence threshold, all else equal. Furthermore, on the right side, the increase in the expected sanction for benign acts—the upper element in the right box—is computed on a different base because benign acts are investigated only when harmful acts are committed. The impact of this difference is hard to assess; one might assume that in many instances there would be far fewer opportunities for mistaken identification to occur, but that will depend on the nature of the benign acts in question and the targeting precision of investigators versus monitors or auditors.

Although this analysis suggests that lower evidence thresholds seem optimal when enforcement is by investigation, the results are not really that comparable across enforcement methods. Each occurs in entirely different situations, in which different sorts of benign acts might be confused for harmful ones, and with different relative likelihoods. And the method of identification differs as well, as may the nature of evidence generated. All that can be concluded with confidence is that the determinants differ in important ways across methods.

Next, examine the difference between the determinants of the optimal evidence threshold and of the preponderance rule (and others like it) when enforcement is by investigation. The two added deterrence benefits have no analogue in the determinants of rules based on ex post likelihoods because, as explored at length in Part II, they do not in any direct way depend on deterrence effects. Recall the observation that the optimal evidence threshold depends on how behavior changes as a consequence of adjusting the threshold, whereas the preponderance rule depends on the aggregate level of harmful and benign acts at a given threshold because these aggregates influence the flow of acts into adjudication. In the present setting, the determinants of the preponderance rule differ somewhat from what they were before,[175] but not in

175. The difference is that the frequency of harmful acts—a key component of the upper element in the left box in Figure 5—no longer enters into the formulation because it influences both sides of the equation in the same way and thus cancels. The reason is that, as explained in the text, commission of each harmful act triggers an investigation, perhaps with some probability; then, conditional on an investigation, there is some likelihood that the

ways that influence the conclusion that they are fundamentally distinct from those for the optimal evidence threshold. Once again, the preponderance rule and others based on the ex post likelihood that the individual before the tribunal committed a harmful act are not even serious proxies for effects on social welfare and thus in no way suggest what evidence threshold is optimal.

Finally, it is worth elaborating on the deterrence effect of a changed evidence threshold with regard to those who commit benign acts. It was already noted in Section II.C that raising the evidence threshold could actually increase the frequency of the mistaken imposition of sanctions. When enforcement is by investigation, this possibility also arises, but in a different manner. Raising the evidence threshold, to be sure, makes it less likely that a given individual who both commits a benign act and ends up before a tribunal will be sanctioned. However, it may also notably increase the number of such individuals who come before the tribunal precisely because the higher threshold reduces deterrence and thus triggers more investigations. Accordingly, individuals committing benign acts may fare worse under a higher evidence threshold.[176] This effect is hardly certain, and it is difficult to say how likely it may be, but it appears to be entirely plausible, not merely conceivable. We therefore have another instance in which we can be misled by an ex post perspective that concentrates on an individual imagined to be before a tribunal and ignores how the evidence threshold influences behavior and thus who comes before tribunals in the first instance.

individual committing the harmful act is identified and some other likelihood that an individual who committed a benign act is identified. Changing the number of harmful acts proportionally changes the flow of both groups into adjudication.

176. *See* Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399, 413 (1973); *cf.* KAPLOW & SHAVELL, *supra* note 2, at 336-38 (observing how high sanctions may similarly benefit the innocent through greater deterrence). Note that in such a scenario innocent individuals would also fare worse ex ante, which means that chilling could rise despite the higher evidence threshold. This result contrasts with that under monitoring (or auditing) examined previously, because there the only way more individuals who commit benign acts might end up being sanctioned despite a tougher proof requirement is on account of reduced chilling, which is the source of the increased number of benign acts entering adjudication. Furthermore, observe that, in the criminal setting, there would be two senses in which innocent individuals may be worse off with a higher evidence threshold: they suffer more as crime victims due to reduced deterrence (the familiar point), and they may more often be mistakenly subject to both trial and conviction, for the reason just given in the text.

C. Regulation of Future Conduct

This Section shifts from the focus of the rest of the Article to a fundamentally different means of legal regulation: that which dictates future conduct. Until now, the analysis has concentrated on how the evidence threshold, by influencing expected sanctions for harmful and for benign acts, influences ex ante behavior. The central welfare effects of the evidence threshold have involved the social benefits and costs of these behavioral consequences: deterrence and chilling. The primary significance of the imposition of sanctions was taken to be these ex ante behavioral effects. Secondarily, but sometimes importantly, it was recognized that sanctions are themselves costly, in which case it is socially desirable, ceteris paribus, to minimize their imposition.

The central purpose of some aspects of the legal system, by contrast, is to regulate future conduct. Many forms of licensing—of drivers, certain professionals (doctors, lawyers), and facilities (restaurants, nuclear power plants)—are designed to permit or prohibit future activity based on whether it is likely to be socially beneficial or detrimental.[177] Zoning regulations control what structures may be built or how they may be used (residential versus commercial or industrial). Competition regimes may require large mergers to be approved in advance, which authorization is withheld if anticompetitive effects are expected to be significant. And injunctions limit parties' future activities.[178]

Because the central effects of the regulation of future conduct differ (in the pure case, entirely) from those examined thus far, Subsection IV.C.1 presents a fresh analysis of how the optimal evidence threshold should be determined in this setting.[179] The problem is analytically simpler than those considered

177. Some licensing is more akin to registration, in many cases to alert authorities about who must be subject to inspection or to document the payment of a required fee.

178. Some of these examples may be less different than meets the eye. Notably, a party bound by an injunction—or, more often, one who might be threatened with suit for an injunction—may settle with the plaintiff, making a payment in return for future freedom of action. The overall effect is more akin to a suit for damages, the prospect of which influences behavior as before.

179. This distinction also proved central—for similar reasons—in my prior work on accuracy in adjudication. *See* Kaplow, *supra* note 102, at 369-81. As in that case, the present framework looks more akin to traditional cost-benefit analysis. *See id.* at 380. Standard cost-benefit analysis would look more like a comparison of the bottom elements in Figure 9, below, supposing certainty. In cost-benefit analysis, uncertainty is often analyzed using decision trees that, when collapsed into an algebraic expression, would resemble the complete formulation in Figure 9.

previously; in fact, the framework is essentially the same as that applicable to medical decisionmaking and hence is familiar.

Subsection IV.C.2 compares the determinants of the optimal evidence threshold for this mode of enforcement with those for the preponderance of the evidence rule and others based on the ex post likelihood that the individual before the tribunal is of the harmful type. On one hand, the two formulations are much more similar than was true in any of the prior cases—where there was almost no resemblance whatsoever. This greater commonality is, on reflection, to be expected because, as emphasized in Part II, conventional conceptions seem to arise from an ex post perspective that takes as given the cases that come before the tribunal,[180] and this stance is appropriate when the legal policy question concerns what future conduct should be permitted rather than what rule for imposing sanctions best influences ex ante behavior in anticipation thereof. On the other hand, in the present context there are still central determinants of the optimal evidence threshold that are omitted from the formulation of the preponderance rule and others of its kind. Moreover, it will be apparent that this divergence means that the optimal threshold will often differ markedly from the preponderance rule (or any other rule of that type), and that the differences can involve an optimal evidence threshold that is much higher or far lower.

Subsections IV.C.1 and IV.C.2 examine a simple, pure form of adjudication aimed at regulating future conduct. In many actual settings, however, such adjudication has a mixed nature, often influencing ex ante behavior as well. One important instance is imprisonment, which has deterrent (and chilling) effects as well as an incapacitation effect via the direct regulation (limitation) of imprisoned individuals' future activities.[181] Likewise, license revocations are often both penalties for past misbehavior and limitations on future activity. Accordingly, Subsection IV.C.3 briefly analyzes how evidence thresholds should be determined in mixed cases, that is, those in which adjudication has important ex ante effects as well as a significant influence on future conduct.[182]

180. As emphasized at a number of points, however, all such rules depend on the flow of cases into adjudication, which in turn depends on ex ante effects. Hence, it is not so much that the rules themselves ignore ex ante behavior but that the analysts who find them appealing tend to adopt an outlook that takes the flow of cases as given.

181. The latter was briefly discussed in note 155 with regard to how differential costs of sanctions on harmful and benign acts influence the analysis of the optimal evidence threshold.

182. That Subsection does not focus on another sort of mixing, noted previously, between different modes of enforcement. For example, there may be both environmental licensing of facilities and also subsequent enforcement to ensure compliance, including monitoring, auditing (inspections), and investigation (such as when contaminants are discovered in groundwater).

1. Optimal Evidence Threshold

For ease of exposition and of comparability with the analysis in Parts I and II, suppose that there are two types of actors, those whose future acts are of the harmful type and those whose future acts are benign—which, again, means that the former involve some level of external social harm and the latter do not. Also as before, individuals in each group vary in the private benefits that their acts would generate.[183]

The tribunal must decide whether to permit the future conduct or to prohibit it.[184] If the individual is given permission, the act—harmful or benign as the case may be—will be committed, generating the private benefits and social harm (if applicable), as just described. If prohibited,[185] no act will be committed, and both private benefits and social harm are taken to be zero.[186]

---

183. This description contains a number of simplifications, including that harmful acts generate a given degree of harm, that benign acts generate no harm, and (as will become clear) that the tribunal does not know individuals' private benefits. None of these assumptions greatly affect the core ideas, in large part because one can think of differences the tribunal can observe as constituting different classes of cases, each subject to its own evidence threshold (and, for example, harm can be understood as the average harm in a class of cases in which the tribunal cannot distinguish the level of harm). Another important assumption is that the magnitudes of different types of individuals' private benefits and of the social harm are independent of the strength of the evidence that the tribunal observes. *See infra* note 190.

Furthermore, it is supposed that individuals' private benefits are positive, so they wish to apply for permission. Individuals whose expected private benefits are negative would not act in any case and thus are irrelevant for present purposes. If there was an application cost—whether a fee or a cost of preparation and appearance—then only individuals with private benefits sufficiently high (not merely above the total application cost, because the private benefit must be discounted to take into account the probability that the application will be denied) would apply and come before tribunals. In that case, the average benefit levels noted below would need to take this truncation into account; moreover, reducing the evidence threshold would also deter some applications (for both harmful and benign acts) and also influence the total application (administrative) costs incurred. These considerations would also be relevant under a complete analysis but, for simplicity, are omitted here.

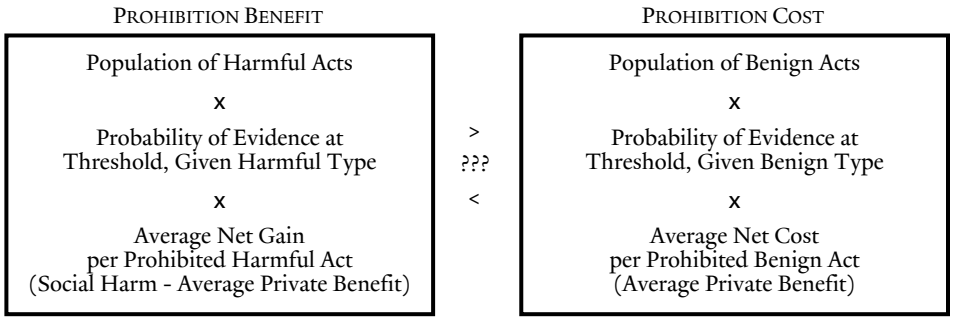
184. In the present analysis, the default is immaterial, and it may as well be taken to be prohibition, with ties (*see supra* note 34) resulting in prohibition. For some acts, such as initial permission to drive, this is the typical legal default, whereas with others, such as the right to go about one's life rather than be civilly committed, the default is permission.

185. In essence, there is an implicit assumption of perfect and costless enforcement of the prohibition. In many contexts, this depiction is not a plausible approximation, and one would have to consider as well enforcement by the sorts of means considered previously, which would be the sort of mixed case discussed in note 198.

186. Some who address these sorts of problems assign nonzero costs or benefits to the prohibition outcomes for the two types of acts. *Cf.* Arkes & Mellers, *supra* note 78, at 627-30 (assigning values to all four outcomes but presenting analysis that depends on only the two

To determine what evidence threshold is optimal in this setting, we can again (as when introducing Figure 4 in Section I.B) inquire into the effects of a slight reduction in the threshold from any given level. The relevant analysis for this case is depicted in Figure 9.

Figure 9.
EFFECTS OF A REDUCTION IN THE EVIDENCE THRESHOLD: FUTURE CONDUCT



The benefit of a lower threshold is that additional acts that would turn out to be harmful are prohibited, and the cost is that additional benign acts, which would generate private benefits at no social cost, are forgone. If the former exceeds the latter, a lower threshold is socially beneficial; if the former is smaller, a higher threshold would be beneficial; and the optimal evidence threshold is characterized by the incremental benefit and cost of any marginal adjustment being in balance.

Observe that the elements in Figure 9 appear, roughly speaking, to be a combination of those from Figure 4, for the optimal evidence threshold in our base case focusing on ex ante behavior (the lower elements being particularly similar), and from Figure 5, for the preponderance rule in that case (particularly regarding the upper and middle elements). To see the similarities and differences more precisely, consider each component in turn. (Explicit comparison to Figure 5 is deferred to Subsection IV.C.2.)

The prohibition benefit—in the left box—is the product of the number of harmful acts that would be prohibited by a lower threshold and the social gain

differences; note that their analysis is inapposite in the typical setting they contemplate, due to the endogeneity of behavior). However, for many purposes, all that matters for each type of act is the difference between the outcome with permission and prohibition, and the benefits and harm described in the text can be taken to be these differences. After all, any notion of gain or loss is relative to some benchmark, and here it is convenient to take as the benchmark (zero point) the consequences when each type of act is not undertaken.

per prohibited harmful act. The former has two elements. First, we wish to know how many individuals would seek permission to commit acts that would be harmful, that is, the population of such acts. Second, of those acts, the portion newly prohibited as a consequence of the threshold reduction will be given by the change in the probability that harmful acts are prohibited. This change was the key component of the top element in the left box in Figure 4 as well, and is given by the magnitude of the slope of the $P_{HARMFUL}(x^T)$ curve in Figure 2 (which indicates the rate of change in that probability as the threshold x^T is changed)—or, equivalently, by the height of the probability distribution curve for the harmful act in Figure 1.[187] Reducing the threshold means that cases with evidence strength extremely close to the thresholds (just below the old one but just above the new one) will be prohibited rather than permitted.[188] Multiplying this factor by the relevant population of harmful acts gives the number of acts newly prohibited.

The number of harmful acts prohibited must then be multiplied by the average net social gain per prohibited act. Much as in Figure 4, this net gain is the social harm avoided minus the private benefit forgone. There is, however, a difference regarding the latter. When harm reduction was accomplished by deterrence—the effect presented in Figure 4—the forgone private benefit is that of the marginal individual just deterred (and hence is equal, as explained in Section I.B, to the prevailing expected sanction for harmful acts). Here, it is not the case that some individuals are deterred, with them deciding who that will be. Instead, among all individuals with harmful acts, the tribunal, as a consequence of a lower evidence threshold, will prohibit essentially a random subset: those whose cases happen to be associated with evidence just at the threshold.[189] Accordingly, the forgone private benefit will be the average private benefit for such acts, which are taken to be equally likely to be associated with evidence at any particular level—the evidence likelihoods, as

187. Abstractly, the same curves and the same analysis as before are appropriate. Practically, the shapes of the curves could, of course, be entirely different because here we are considering a different kind of factfinding: a prospective determination of which sorts of individuals would cause harm if permitted to act, rather than a retrospective analysis of whether acts already committed are harmful rather than benign.

188. For convenience, the relevant probability is referred to as that *at* the threshold, when in fact we are considering two thresholds, the new one slightly below the original one. As "slight" becomes very slight (strictly speaking, with a derivative, infinitesimal), any difference in probability levels in the pertinent regions vanishes.

189. This difference explains why the left box is labeled a prohibition benefit rather than a deterrence benefit, and the right box a prohibition cost rather than a chilling cost.

throughout this Article, being influenced solely by whether the act is of the harmful or benign type.[190]

The determination of the prohibition cost is qualitatively the same. We multiply the population size for benign acts by the corresponding probability of evidence at the threshold to determine the number of benign acts that will be prohibited as a consequence of the reduced threshold. This number is multiplied by the average social cost per prohibited benign act, which in turn equals the average private benefit of such acts—again, the average being pertinent rather than the marginal benefit that was relevant when considering chilling costs.

Return now to the characterization of the optimal evidence threshold. There is an important respect in which the analysis here, corresponding to Figure 9, is more straightforward than that needed for Figure 4 because, previously, where behavior was endogenous, many of our components were moving targets. Here, the population fractions—the upper elements in the boxes in Figure 9—are constant. Likewise, the values in the lower elements are also taken to be constant. Accordingly, the first and third elements in each box can be multiplied. To determine the optimal evidence threshold, we need to choose the level of x^T such that the middle elements will offset the others. For example, if the product of the first and third elements in the left box, for the prohibition benefit, is twice the product of the corresponding elements in the right box, for the prohibition cost, then at the optimal threshold the probability from the middle element of the left box will necessarily be half that for the middle element of the right box. We can return to Figure 1 and determine the optimal x^T as that which has this property; it would be an x^T somewhat to the left of center in the figure.[191]

The formulation for this case involving the regulation of future conduct is also in accord with intuition. Consider each of the three elements in the boxes: If the population of harmful acts is larger than that of benign acts, a lower evidence threshold is socially advantageous because lowering the threshold prohibits more harmful acts than benign ones, all else being equal. At the current level of the evidence threshold, the greater the probability of evidence

190. If the level of harm and of private benefits was correlated with evidence strength, then the social harm and average private benefit would each be conditional on the strength of the evidence at the threshold in question.

191. One could also determine this x^T from Figure 2, it being the point where the slope of the $P_{BENIGN}(x^T)$ curve is twice that of the $P_{HARMFUL}(x^T)$ curve. And it can be determined implicitly from Figure 3, as the point where the slope of the bowed curve for $P_{BENIGN}(x^T)$ as a function of $P_{HARMFUL}(x^T)$ equals two. Note further that the posited shape of this curve, which is discussed in Subsection I.A.2, implies that this point will be unique.

being at (very near) the threshold for harmful acts than for benign ones—that is, the more strongly evidence at the current threshold signals that the act before the tribunal is harmful rather than benign—the better it is to lower the threshold. Finally, the greater the net social gain per prohibited harmful act (which depends both on the external social harm and the private benefit from such an act) relative to the cost per prohibited benign act, the more desirable is a lower evidence threshold. And the overall optimal threshold reflects a simple product of these three considerations.

2. Conventional Conceptions Compared

For concreteness, let us compare the determination of the optimal evidence threshold with that for the evidence threshold which satisfies the preponderance of the evidence rule. Just as in Section II.A, we could readily extend the analysis to any ex post likelihood (such as 75%, which some loosely associate with a clear and convincing requirement) by incorporating a multiplicative factor (in this example, a factor of 3 on the right side).

Unlike in Part II, we do not need to present a new diagram because the connection between the preponderance rule and the optimal evidence threshold is much closer analytically. Indeed, if one modifies Figure 9 by omitting the lower elements in each box—and inserts an equals sign between the two boxes—one has a formulation for the preponderance rule.[192] To see this, we can restate this basic proof conception and relate it to the elements in the figure.

Under the preponderance rule, the appropriate evidence threshold is that at which the probability that the act is of the harmful rather than the benign type is 50%, which is to say that it is equally likely that the act is of each type. To apply this notion, we need to determine the two likelihoods for the present setting.

As before, if the two types of acts were equally likely to come before the tribunal, the appropriate evidence threshold would be that at which the probabilities of evidence strength being at the threshold were the same for harmful and for benign acts. Recall that this is the point at which the two

192. It is worth reflecting on why there is this important similarity in the present context that was absent before. Here, the truth of the matter—whether the individual before the tribunal is one associated with the harmful rather than the benign type of act—is directly relevant because what the tribunal proposes to do is determine whether the individual should be permitted to act in the future. Before, the effect of the tribunal's decision was on the level of expected sanctions for the two types of behavior, the prospect of which influenced those behaviors. What mattered in a given case was not the truth of the matter but instead how the prospect of one decision versus the other influenced ex ante incentives for the two types of individuals' decisions.

probability curves in Figure 1 cross—the probability being greater for harmful acts to the right of this point and greater for benign acts to the left of this point.

Also as before, these probabilities—the second elements in each of the boxes in Figure 5 and in Figure 9—are only part of the story because, in general, the two types of acts do not come before the tribunal with equal frequency. In some instances one type of act will arise more often (perhaps much more often), and in other instances it will be the other. These frequencies were given by the first elements in Figure 5 and are now indicated by the first elements in Figure 9.[193] Regarding the latter, we do not strictly speaking have the frequencies, as we did in Figure 5; instead we have the total population (quantity) of each type of act, but this difference is inconsequential for present purposes. If, say, there were 70 harmful acts and 30 benign ones, these will be the populations corresponding to the upper elements in Figure 9; whereas if we instead used frequencies in the top row, in this example we would have 70% harmful acts and 30% benign ones. Using population totals or fractions does not affect the direction of inequality or the evidence threshold at which the two sides are equal.[194]

Recall that the presence of the first elements in the figure is highly consequential. In some instances, most possible acts will be harmful, in which case a very low evidence threshold would implement the preponderance rule, whereas in others most will be benign, in which case a very high threshold would be indicated.[195]

193. Note that the complications examined in detail in Part II, especially in Subsection II.C.1, do not arise in the present setting because behavior is no longer taken to be endogenous. That is, the population sizes or fractions are taken here to be fixed. If we introduced endogeneity through a filing decision—as discussed in note 183—then these issues would return, as they also do in the mixed cases examined in Subsection IV.C.3.

194. For example, if we considered population figures that do not sum to 100 (say, 140 and 60), then the upper elements in the two boxes in Figure 9 would differ by precisely the same factor (in this instance, they would each double), and multiplying both sides of an inequality or equation by the same positive number does not affect the relationship.

195. As discussed at the end of Subsection II.C.1, the use of Bayesian priors—which here correspond to the population fractions (themselves proportional to the upper elements in the boxes in Figure 9)—is controversial. There it was noted that the controversy may be avoided because such priors are immaterial in determining the optimal (welfare-maximizing) threshold in that setting. Here, however, these priors, as just explained, are highly consequential: they could tip the optimal evidence threshold in either direction, and substantially so. Hence, it would be socially detrimental if queasiness about the use of Bayesian priors influenced the determination of decision rules for the regulation of future conduct. The problem is equivalent to that in the medical decisionmaking context, noted below in the text, where it would be unimaginable in interpreting tests and other diagnostic

At this point, we can see that there is one qualitative difference between the formulation for the optimal evidence threshold and that for the preponderance rule: the former has an additional pair of elements, the lower ones in Figure 9, that are omitted from the latter. This divergence mirrors one of the key distinctions considered in Subsection II.B.1, and it again can be extremely important.

To appreciate its significance, begin with the evidence threshold that implements the preponderance rule: it by definition is one for which the product of the first two elements in the left box in Figure 9 equals the corresponding product for the right box. Now we can examine whether and how the third elements may differ. Even supposing that the private benefits for the two types of acts are the same (which in general they need not be), we note that they enter the two sides with opposite signs. That is, the average private benefit of a harmful act reduces the benefit side of the calculus but increases the cost side of the calculus. This opposition is hardly surprising because individuals' private benefits are social costs of prohibition for both types of acts. Moreover, the left, benefit side has the external social harm, which is avoided by the prohibition of harmful acts, and the right, cost side has no corresponding component. Hence, the determinants of the two sides are entirely different.

Only by coincidence could we expect the third elements to be equal.[196] They could differ in either direction, and the differences could be small or large. It follows that the ex post likelihood at the optimal threshold could be anything: it would be very high when acts posed little harm relative to private benefits, and it would be very low when acts posed great harm. To illustrate, the degree to which one administering a driving test should be convinced that the applicant would generally be safe should be quite different for an ordinary driver's license than for a license to operate vehicles carrying hazardous substances.

evidence to ignore base rates (which, as here, can vary radically across contexts) in prescribing treatment.

196. Suppose that one both adopts an ex post perspective that takes behavior as given (which this Section indicates is more nearly appropriate regarding the regulation of future conduct) and then further stipulates that error costs are equal—essentially an ad hoc notion that, from some unstated normative perspective, all errors are of equal importance, a view that is widely endorsed by courts and commentators alike in the civil setting (*see* sources quoted *supra* note 7). Under these assumptions, the social objective reduces to the minimization of the total number of errors, which, as is familiar (when one implicitly takes the flow of cases into adjudication as given), implies the optimality of the preponderance rule. *See supra* note 111; *see also* John Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065, 1071-72 (1968) (noting that a preponderance standard supposes that the two types of errors have equal cost); Posner, *supra* note 176, at 408 (same).

Moreover, this point is regarded as obvious in other settings, like medical decisionmaking. If a medical procedure is very expensive and highly risky, it is sensible to employ it only when tests indicate that the patient is almost certain to be a type who would benefit (some cancer treatments). But if a therapy is inexpensive and poses almost no risk of serious side effects, it might optimally be used, perhaps prophylactically, even on patients who only have a small likelihood of any benefit (some vitamins).

To summarize, the difference between the preponderance rule (or others like it) and the formulation for the optimal evidence threshold is, qualitatively, quite less in the present setting than in those centrally concerned with ex ante behavior. Instead of an almost complete disconnect, we have a substantial overlap of relevant factors. Nevertheless, the preponderance rule omits considerations of harm and benefit, which are of critical importance if one is concerned with the consequences of adjudication for social welfare.

Interestingly, in many settings in which decisions are made about the permission or prohibition of future conduct—which often is not done through conventional adjudication—we see implicit or explicit use of cost-benefit analysis that is much like the analysis in Subsection IV.C.1 for the optimal evidence threshold rather than deployment of a preponderance of the evidence rule (or others like it). Many licensing and zoning decisions are made by regulatory bodies that undertake an inclusive analysis that attends to social harm and private benefits. New doctors and new drugs must pass certification tests, and the minimum threshold is not set based on a pure likelihood calculus, one presumes. Even for injunctions determined by courts, factors include considerations of harm and benefit, not just likelihoods.[197] It is notable that when the likelihood test is less distant from a sensible formulation, the legal system tends to employ an explicit welfare-based analysis, whereas when it is not even close, the legal system tends to use ex post likelihood rules. Perhaps this difference reflects that, when the deficiencies of a likelihood-based approach are simple and glaring, its appeal evaporates.

197. *See, e.g.*, eBay Inc. v. MercExchange, L.L.C., 547 U.S. 388, 391 (2006) ("[A] plaintiff seeking a permanent injunction must satisfy a four-factor test before a court may grant such relief. A plaintiff must demonstrate: (1) that it has suffered an irreparable injury; (2) that remedies available at law, such as monetary damages, are inadequate to compensate for that injury; (3) that, considering the balance of hardships between the plaintiff and defendant, a remedy in equity is warranted; and (4) that the public interest would not be disserved by a permanent injunction."). Regarding preliminary injunctions, where the concern for error is at the fore, it is standard to take into account the costs to the two parties of different outcomes as well as the probabilities that one or another result on the merits is correct. *See* POSNER, *supra* note 7, at 760-61; John Leubsdorf, *The Standard for Preliminary Injunctions*, 91 HARV. L. REV. 525 (1978).

3. Mixed Cases

The analysis to this point of the regulation of future conduct examines a simple, pure scenario in which individuals—certain numbers of whom are associated with harmful and with benign acts—seek permission to perform future acts and the legal system either grants or denies permission, which determines which acts are committed. This Subsection considers mixed cases[198] in which the prospect of prohibitions influences ex ante behavior as well.[199] Examples already given were imprisonment—the prospect of which deters (and chills) acts and the imposition of which governs future conduct for a time—and license revocation, such as for drunk driving—the prospect of which deters (and chills) certain driving behavior and the imposition of which

198. Another type of mixed case arises when a prohibition is not self-executing. Then one is, from that point on, back to the situation examined in the rest of this Article. Indeed, that analysis assumed that all harmful acts were prohibited to begin with, and the problem is that individuals nevertheless commit harmful acts when private benefits exceed the expected sanction. The existence of prior proceedings may still matter. One possibility is that the prior determination may be viewed as decisive with regard to aspects of liability, so the only question may be whether a person acted. For example, the offense of driving without a license may require proof that the person denied a license was driving, but there is no further determination needed regarding whether the act is regarded to be harmful. (For some regulations, violations would be sufficiently obvious that the rules may be regarded as self-enforcing. For example, a utility cannot very well secretly operate an unlicensed nuclear power plant. Nor is a major pharmaceutical company likely to sell a drug prohibited by the regulators, although obviously the sale of some prohibited drugs is widespread and enforcement is very costly.) In other settings, there may be little difference. If what is illegal is not operating a plant but discharging a substance (where permission might sometimes be granted), one will need to determine whether harmful substances were discharged, by whom, and so forth.

199. *Cf.* Kaplow, *supra* note 102, at 379-81 (examining how mixed cases affect the analysis of accuracy in adjudication). Similar analysis, with some of the effects reversed, is applicable to determination of eligibility for future rewards, which could include intellectual property rights (the granting of a patent—which, note, also constitutes a conditional prohibition on others' behavior) or transfer payments (for example, disability benefits determinations). The latter case also brings to mind the compensatory function of some private litigation: that is, lawsuit outcomes may matter not only on account of the provision of ex ante incentives but also because compensation may be independently beneficial (or detrimental) on account of parties' risk aversion. Although the present Section focuses on future conduct, the analytical structure really refers to any consequence that follows directly from the outcome of adjudication, by contrast to effects due to the anticipation of adjudication outcomes. *Cf. id.* at 369-78 (addressing the value of accuracy in determining future entitlements); *id.* at 383-86 (discussing the accuracy of compensation); Louis Kaplow, *Optimal Insurance Contracts When Establishing the Amount of Loss Is Costly*, 19 GENEVA PAPERS ON RISK & INS. THEORY 139 (1994) (analyzing the value of accuracy in providing compensation through insurance).

prevents future driving by the individual. Clearly, many, although not nearly all, denials of permission influence ex ante behavior of various sorts, including myriad types of investment: putting together a development project, which requires permission from a zoning board to go forward;[200] studying to become a doctor, which requires obtaining a medical license; or research and development in the pharmaceutical industry, which requires drug approval.[201]

The optimal evidence threshold in these mixed cases will reflect a weighted average of the two prior calculuses. In conducting the balance of benefits and costs, we can sum the results in the left boxes in Figures 4 and 9 and the results in the right boxes, and then compare the two sums. Because reducing the evidence threshold has both sets of effects, both of the benefits (deterrence and the social gains from the prohibition of future harmful acts) and both of the costs (chilling and the social costs from the prohibition of future benign acts) are pertinent. And clearly, depending on which sorts of effects are largest, the optimal threshold may more nearly resemble one from the former pure case or from the latter one.

D. Selection of Cases for Adjudication

Return to the core setting in which adjudication examines past acts to determine whether sanctions should be imposed. Most of the analysis of this Article takes as given that some portion of harmful acts and of benign acts will come before tribunals. The exception is Section III.A, which examines how optimally to choose the level of enforcement effort in conjunction with setting an evidence threshold. However, with both public and private enforcement, the process by which cases are identified and brought forward is not mechanical but rather depends on the incentives of government enforcers and private litigants. Even more relevant for present purposes, the manner in which cases are selected for adjudication may bear on the optimal evidence threshold, in part because the evidence threshold itself may influence the flow of cases into adjudication.[202] This

200. An important analogous example involves merger approval. Not only is the application process nontrivially costly, but since there is a serious prospect of denial for certain mergers, the ex ante incentives to find possible merger partners, negotiate deals, arrange financing, and so forth will be reduced, for both harmful and benign mergers, in degrees that depend on the subsequent standard for approval.

201. This phenomenon includes but is not limited to the ex ante behavior required to file the application itself—a subject discussed briefly in note 183. Because the analysis is conceptually quite similar, no sharp distinction need be drawn.

202. As mentioned at some points below, the evidence threshold also influences parties' incentives to generate evidence.

complex relationship raises myriad issues, only a handful of which are touched upon here.

1. Public Enforcement

Most of the discussion has implicitly or explicitly contemplated public enforcers, such as when the government decides how to deploy a police force; conduct tax audits or environmental, health, and safety inspections; and prosecute cases. However, even for a given budget and set of rules guiding government officials in these tasks, one must take into account the incentives they face. Agents' rewards may be variously affected—perhaps directly but often more subtly on account of norms, promotion and retention criteria, and elections—by many factors, including: the number or portion of investigations or audits that generate prosecutions, the quantity or percentage of successful prosecutions, the quantity of fines or other sanctions imposed, the rate of crime or other infractions and changes therein over time, complaints about abuse, and embarrassment about exposed error. Some of these are influenced by the proof burden to which cases will be subject, and some may in any event be relevant to how it should be set. These phenomena are likely to be important because such incentives determine both the cases that are pursued and the evidence that is developed.[203]

Suppose, for example, that the evidence threshold is raised for a set of violations. If one assumes that enforcers are most heavily influenced by the fraction of cases that ultimately are prosecuted successfully, they might pursue fewer cases, putting more effort into those that remain. The consequence might be that the increase in the evidence threshold would reduce deterrence by even more than if such a response was ignored because prospective violators

203. Prior discussions of how proof burdens may affect prosecutors' and defendants' efforts in litigation include Thomas J. Miceli, *Optimal Prosecution of Defendants Whose Guilt Is Uncertain*, 6 J. L. ECON. & ORG. 189 (1990); Daniel L. Rubinfeld & David E.M. Sappington, *Efficient Awards and Standards of Proof in Judicial Proceedings*, 18 RAND J. ECON. 308 (1987); and Okan Yilankaya, *A Model of Evidence Production and Optimal Standard of Proof and Penalty in Criminal Trials*, 35 CAN. J. ECON. 385 (2002). *See also* Tracy Lewis & Michael Poitevin, *Disclosure of Information in Regulatory Proceedings*, 13 J. L. ECON. & ORG. 50 (1997) (analyzing applicants' incentives to present evidence in regulatory proceedings of the sort analyzed in Section IV.C); Derek Pyne, *Can Making It Harder To Convict Criminals Ever Reduce Crime?*, 18 EUR. J. L. & ECON. 191 (2004) (examining a model in which police are rewarded for convictions, regardless of their accuracy, and finding that a higher proof burden may increase their efforts to identify the truly guilty and thereby increase deterrence); Matthew C. Stephenson, *Evidentiary Standards and Information Acquisition in Public Law*, 10 AM. L. & ECON. REV. 351 (2008) (studying the effect of reviewing courts' evidentiary demands on agencies' incentives to acquire information).

care not about the likelihood of being sanctioned conditional on their case being pushed forward but rather on the overall likelihood of sanctions. Furthermore, if the cases dropped are weaker ones and some additional ones are winnowed out when more intensive effort reveals exonerating evidence, the ex ante likelihood of mistaken imposition of sanctions may also fall by even more than they would without this reaction by enforcers. Note also that a more complete analysis needs to account for defendants' and their lawyers' incentives. With a higher burden on the enforcer, the defense might expend less effort because it is easier to prevail, which could partly offset the aforementioned effects; but sometimes the defense might strive harder because some nearly hopeless cases may have become more winnable. It is important to keep in mind in this instance and others that enforcers' behavior can to a degree be controlled, such as in setting agency budgets and adjusting actors' financial and other rewards.

Other effects are possible with different incentives. One set of considerations is particularly important, especially in cases that involve political opponents of a regime or other individuals the government would improperly like to attack through the legal system. In such situations, a high evidence threshold may be especially valuable in light of the government's ability to concentrate significant resources on such cases[204] and possibly attempt to fabricate evidence.[205] Moreover, the resulting chilling effects (stifling of political opposition) if such prosecutions are too easy could have a particularly high social cost.[206] For these reasons, a very tough threshold may be optimal. The problem is that it is hard to know which cases should be so treated, and having to treat all cases (or large sets of them) this way can itself be quite

204. In discussions of government prosecutions, criminal and civil, it is often remarked that the government's resources vastly exceed those of private parties. The comparison, however, is quite misleading because, in general, the government needs to spread its resources across many priorities, only a small fraction of which involves police and prosecution, and those resources in turn must cover all manner of violations by the entire population. In many settings—notably some white collar crime—the government is often significantly outspent by private parties. The important caveat, which the text emphasizes, is that the government has great discretion over how it allocates resources across cases. It can choose to concentrate them on particular individuals or firms—or threaten to do so, perhaps to great effect. *See* POSNER, *supra* note 7, at 828-29 (explaining that prosecutors can threaten to expend sizable resources at trial in order to induce favorable plea bargaining outcomes).

205. Relatedly, a high burden on the prosecution makes it more difficult for the government to impose large defense costs on its targets because they can succeed in many instances with more modest effort.

206. Another class of abuse concerns prosecutions motivated by the personal gain of the prosecutor. One seeking promotion, reelection, higher office, or fame might pursue high-profile defendants who may well not be political opponents or even be politically associated.

troublesome. The difficulty is especially great because another priority of many governments, including those from whom abuse may most be feared, is combating corruption, which requires successful prosecution of behaviors in settings that may be observationally similar to those with a high potential for improper enforcement activity.

This argument for a particularly high evidence threshold may help to explain the requirement of proof beyond a reasonable doubt in U.S. criminal cases as well as its constitutional status, along with other procedural protections distinctive to the criminal sphere,[207] many of which seem motivated by concerns about abuse of power rather than cost-benefit calculations for routine policy-making by benevolent officials.[208] Bolstering this view is the commonplace advancement of a preponderance rule for civil cases on the ground that the special concerns in the criminal context are absent, so society has no particular reason to tilt the legal system in favor of one outcome versus the other.[209] Note, however, that all manner of abuse can arise whenever the government (rather than a private plaintiff) is the prosecuting party, which arguably calls for a higher evidence threshold regardless of whether the legal action involves a criminal proceeding. Moreover, when targeted defendants are corporations (or other organizations) and the main penalties are fines or possible reputational costs, civil proceedings may be quite a good substitute in advancing the objectives of overly zealous government officials.[210]

207. Another rationale is that criminal cases may involve the deprivation of liberty, although the analysis in Subsection III.B.2 shows how higher social sanction costs actually have ambiguous implications for the optimal evidence threshold.

208. *See, e.g.*, Keith N. Hylton & Vikramaditya Khanna, *A Public Choice Theory of Criminal Procedure*, 15 SUP. CT. ECON. REV. 61 (2007). Observe further that if this set of protections is largely effective—that is, in deterring government abuse, although possibly at the expense of higher ordinary crime—examination of trial outcomes in equilibrium may give the impression that the proof standard is too high, illustrating the analysis in Section II.C that indicates how an ex post view focused on outcomes is highly misleading regarding whether the evidence threshold is set optimally.

209. *See* sources quoted *supra* note 7. Note further that the Supreme Court's analysis of proof standards in civil cases employs the balancing test of *Mathews v. Eldridge*, 424 U.S. 319 (1976). *See, e.g.*, Santosky v. Kramer, 455 U.S. 745, 754 (1982); *see also* Kaplow, *supra* note 102, at 373-78 (suggesting respects in which this cost-benefit test is confused). With regard to the termination of parental rights on grounds of neglect and the use of involuntary civil commitment—where the stakes involve liberty—the Supreme Court has applied this test to require proof by clear and convincing evidence (but not proof beyond a reasonable doubt). *See Santosky*, 455 U.S. 745; Addington v. Texas, 441 U.S. 418 (1979). *But see supra* note 55 (criticizing the assumption that higher stakes generally favor a more stringent evidence threshold).

210. Finally, as the next Subsection suggests, it is hardly obvious that we should generally be more concerned about overzealousness by government enforcers than by private ones. In

2. Private Enforcement[211]

In civil cases with private plaintiffs, the primary determinant of suit will be whether a prospective plaintiff's expected recovery exceeds its expected litigation costs.[212] Furthermore, plaintiffs' expected recoveries depend on the evidence threshold. In particular, higher thresholds discourage suit, which reduces deterrence and chilling,[213] whereas lower thresholds encourage suit and thereby heighten deterrence and chilling.[214]

many routine settings, the government may tend to be too lax, whereas profit-motivated private parties might be highly creative and strategic in bringing cases that have socially detrimental effects. *See, e.g.*, Alan Devlin & Michael Jacobs, *Antitrust Error*, 52 WM. & MARY L. REV. 75, 127-30 (2010) (advocating that less weight be placed on concerns for false findings of liability in antitrust cases brought by government agencies rather than by private plaintiffs).

211. Prior economic analysis of the burden of proof and parties' incentives to expend resources on lawsuits and on the presentation of evidence includes Antonio E. Bernardo, Eric Talley & Ivo Welch, *A Theory of Legal Presumptions*, 16 J. L. ECON. & ORG. 1 (2000); Demougin & Fluet, *Rules of Proof*, *supra* note 102; Hay & Spier, *supra* note 5; and Chris William Sanchirico, *A Primary-Activity Approach to Proof Burdens*, 37 J. LEGAL STUD. 273 (2008). *See also* A. Mitchell Polinsky & Steven Shavell, *Legal Error, Litigation, and the Incentive To Obey the Law*, 5 J. L. ECON. & ORG. 99 (1989) (examining the effects of errors on the incentive to sue and discussing how fines and subsidies could be employed to alter this incentive).

212. The calculus is more subtle on account of strategic factors related to asymmetric information and the sequencing of parties' expenditures (which may make some negative expected value suits viable), risk aversion, and other considerations. *See generally* Kathryn E. Spier, *Litigation*, *in* 1 HANDBOOK OF LAW AND ECONOMICS 259, 305-07 (A. Mitchell Polinsky & Steven Shavell eds., 2007) (surveying pertinent literature). These details are not central for present purposes. Relatedly, the present discussion also abstracts from settlement.

213. The text oversimplifies in a manner that depends on the nature of prospective plaintiffs' information. To begin, note that the suggestion is that a higher evidence threshold discourages suit in a particular case, taking as given what a prospective plaintiff knows ex ante (a subject further discussed at the end of this Subsection). Note that, even though suit conditional on, say, a harmful act, may be less likely, the deterrence reduction could raise the total number of suits and thus legal expenditures. Moreover, as deterrence falls, it is possible that individuals who suffer harm and are uncertain (perhaps before discovery) whether their injury was due to improper behavior might be more likely to sue because they will realize that true liability is more likely. In equilibrium, suits cannot be more likely per harmful act, for in that case deterrence would be greater, which would nullify (reverse) the factor encouraging suit, leaving only the direct effect of a higher evidence threshold in deterring suit. *See* Bernardo, Talley & Welch, *supra* note 211 (analyzing a model of the negligence rule in which this inference phenomenon is operative).

214. These effects do tend to favor an intermediate proof standard, for an extremely low one would encourage all manner of suits with little merit (mostly involving benign acts) and an extremely high one would discourage even most meritorious suits (mostly involving harmful acts). However—despite the Goldilocks Principle—knowing that the optimal evidence threshold requires more than nothing and less than infinity hardly suggests that a

As a consequence, recognition that the selection of civil disputes for adjudication depends on the evidence threshold tends to augment the behavioral effects of adjusting the evidence threshold. Hence, in determining what threshold is optimal, such as in the setting examined in Part I, one would need to adjust (raise) one's measurement of the impacts on expected sanctions (the upper elements in Figure 4). Note that it is a priori ambiguous which way this consideration cuts because both the deterrence benefit and the chilling cost of a lower threshold are increased. The direction of the net effect would depend, among other things, on the likelihood that cases at the margin of profitability for prospective plaintiffs are valid.

In Section III.A, it was supposed, by contrast, that the level of enforcement effort could be chosen by the state, along with the evidence threshold; that is, they could be optimized together. The current remarks, and similar ones in Subsection IV.D.1, suggest that this effort level is to a significant degree endogenous—that is, responsive to how the evidence threshold is set. However, in both instances, this story is incomplete. For public enforcement, it remains true that budgets and other policy instruments significantly influence enforcement effort. Moreover, some of the incentives are themselves publically determined: for example, compensation and promotion criteria can be chosen with effects on enforcement effort in mind.

Similar logic extends to private enforcement, although some of the mechanisms are different. For a given evidence threshold, one can encourage or discourage suits through the offering of subsidies or setting of fees, adjusting rules for damages (including through the use of multipliers and decoupling), fee shifting, and other tools.[215] Therefore, one still needs to take into account how setting the evidence threshold affects incentives for suit, but there remains substantial room to adjust enforcement effort separately. Indeed, especially with the private enforcement of public law, fee shifting (sometimes one-way) and damages rules (including statutory minimums) are selectively deployed to influence suit. Also, various procedural rules, such as standards governing dismissals and class actions, may be customized.[216]

50% rule, or anything close to it, is optimal. In this regard, as mentioned in the Introduction, the attraction of the 50% rule probably has much to do with its power as a focal point and the lack of alternative focal points, other than 0% and 100%, which are obviously ruled out.

215. *See, e.g.*, Louis Kaplow, *Shifting Plaintiffs' Fees Versus Increasing Damage Awards*, 24 RAND J. ECON. 625 (1993); Spier, *supra* note 212 (surveying literature on many of these instruments and others).

216. Consider, for example, the Private Securities Litigation Reform Act of 1995, Pub. L. No. 104-67, 109 Stat. 737 (codified as amended in scattered sections of 15 U.S.C.).

Another consideration, mentioned with regard to public prosecutors, is that private litigants' incentives to develop evidence—through investigation, analysis, and trial preparation—are influenced by the evidence threshold. For example, a higher evidence threshold, in addition to discouraging some suits, may encourage plaintiffs to spend more on those that they do bring. In such cases, defendants may spend less, because it is easier to win, or more, because previously unpromising cases now present serious prospects for success. These efforts have a further influence on the deterrence and chilling effects of the legal system and also constitute resource costs in themselves, all of which bears on what evidence threshold is optimal.

Regarding both private and public enforcement, it has also been implicitly assumed that the quality of evidence, however generated, does not become apparent until adjudication itself. At that point, liability is assessed if and only if the strength of the evidence exceeds the stated threshold. Consider, by contrast, the opposite extreme: where parties know before trial exactly how strong the factfinder will perceive the evidence to be and accordingly can predict the outcome perfectly.[217] Then, as long as adjudication involves any cost, public and private enforcers would drop all cases where the evidence strength x failed to exceed the threshold x^T. As a consequence, all adjudicated cases would find in favor of the prosecutor or plaintiff.[218] Note that, in this simplified scenario, none of the *outcomes* would be different; instead, for every defendant victory we would substitute a dropped case.

If, however, one attempted to assess the effectiveness of the legal system based on the outcome of adjudicated cases—a precarious endeavor even without regard to the current point, as developed in Subsection II.C.2—one's view would be skewed even further. There would be no mistaken exonerations

217. The latter conclusion requires the additional assumption that the factfinder always follows instructions; if not, one could suppose further for the purposes of the hypothetical illustration that deviations can also be predicted perfectly.

218. This statement ignores the possibility of settlement, which parties would find advantageous. More broadly, the analysis assumes that the legal rule is formulated as an evidence threshold rather than as an ex post likelihood. As explained in Subsection II.C.1, the relationship between the two notions is complex due to the endogeneity of behavior. The present discussion adds a further complication to that story, for factfinders, in assessing the frequencies with which harmful and benign acts come before tribunals—the upper elements in Figure 5—would now also have to take into account parties' incentives to drop and settle cases. In the scenario in the text, for example, factfinders would know that any case that reached them was one meriting a finding of liability, so they would always do so. But if parties anticipated this, even those with weak cases would not drop them or settle. The ultimate result would be an equilibrium in which a greater variety of cases came before tribunals. Application of the preponderance rule (or others like it) would accordingly be even more difficult than previously suggested.

because all cases with that prospect would be dropped, and there would be no correct exonerations for the same reason, so one hundred percent of individuals who actually committed benign acts but are nevertheless subject to adjudication would be mistakenly subject to sanctions. The system would appear much worse on this dimension when in fact it would be no different; it would merely be saving the resource costs of adjudication in cases that would not assign liability.

Realistically, the truth is somewhere in between what has implicitly been assumed throughout and this extreme hypothetical alternative. That is, parties will to some degree know how a factfinder is likely to assess the evidence in a given case, but there will remain significant uncertainty—due, for example, to incomplete discovery, unpredictable variations in how evidence will be presented (for example, how witnesses will hold up under cross-examination), and diversity in factfinder perceptions. Presumably, in close cases—those in which x is near x^T—prediction will be difficult, but in cases that are not at all close, parties will be nearly certain of the outcome. These differences will, no doubt, significantly affect which cases are pursued and which are settled (including through the use of plea bargaining). And these considerations are also relevant to how the evidence threshold should be set because settlements influence both expected sanctions—for harmful and for benign acts—and the administrative costs of adjudication.

CONCLUSION

The burden of proof is a central feature of adjudication. Nevertheless, the normative question of how high it should be set has received remarkably little attention, particularly in the civil context. This Article begins to fill the vacuum, offering a preliminary conceptual analysis of the subject. It addresses the question of how strong evidence should be in order to assess liability when the objective is the maximization of social welfare.

The initial focus is on a basic law enforcement setting in which some fraction of harmful acts as well as a portion of benign ones will be brought into adjudication. Demanding stronger evidence as a prerequisite to imposing sanctions will dilute deterrence, to some extent undermining the core mission of the legal system: controlling harmful conduct. There is, however, a concomitant benefit: reducing the expected sanctions on benign acts, which decreases the chilling of desirable activity. This tradeoff of deterrence and chilling determines how the optimal evidence threshold should be set. The analysis reveals that it depends on the extent to which raising the evidence threshold reduces the expected sanction for both harmful and benign acts, the number of additional harmful and benign acts committed as a consequence of

such reductions, the magnitude of the benefit produced by each type of act that is no longer discouraged, and the harm caused by the additional harmful acts committed. On reflection, all of the pertinent considerations identified and their interaction in determining an optimal evidence threshold seem natural if the goal is indeed to advance social welfare. Unfortunately, the relevant magnitudes are likely to vary greatly by context and will often be difficult to measure, so the practical problem of determining optimal evidence thresholds is daunting.

This approach is contrasted with conventional conceptions of the burden of proof. Analysis focuses on the preponderance of the evidence rule, which assigns liability when it is more likely than not that the defendant committed the harmful act. It is noted, however, that this rule is qualitatively similar to others that may impose a more stringent requirement—such as demanding clear and convincing evidence or proof beyond a reasonable doubt—because under all such formulations the matter addressed is the likelihood that the individual before the tribunal committed the harmful act.

When one compares the determinants of the evidence threshold under these traditional rules to the determinants of the optimal evidence threshold, it becomes apparent that there is virtually no overlap. Accordingly, neither the preponderance rule nor other likelihood-based rules are even plausible proxies for the evidence threshold that maximizes social welfare. The optimal threshold in a particular context could be far more stringent or much more lax than that implied by any given likelihood-based rule. The core reason for the divergence is that conventional formulations inquire as to the nature of the case before the tribunal, whereas setting an optimal evidence threshold depends on how *changes* in the level of the threshold will influence individuals' ex ante behavior and, moreover, on the welfare consequences of such effects.

The preponderance rule and related formulations also present a number of internal problems. Some are attributable to the fact that the evidence threshold influences behavior, which in turn affects the mix of cases before the tribunal, which itself is a determinant of the likelihood that the factfinder is supposed to assess. This phenomenon renders application of traditional criteria far more complicated than seems to be appreciated by courts or commentators. It also gives rise to counterintuitive possibilities. Notably, one might have supposed that raising the evidence threshold makes it more likely that cases just at the new, elevated threshold involve harmful rather than benign acts because they involve stronger evidence. But this is only part of the story. A heightened threshold also reduces deterrence and the chilling of benign acts, and the latter effect means that more individuals who commit benign acts will come before tribunals. If this effect is sufficiently large—and it is entirely plausible that sometimes it will be—then the net effect of demanding stronger evidence will

be to *reduce* the likelihood that cases just at the new, elevated threshold involve a harmful act rather than a benign one. It is further demonstrated that this feature makes it possible that multiple evidence thresholds will satisfy, say, the preponderance rule, and it is also possible that none will. Likewise, demanding stronger evidence can increase the likelihood that innocent individuals are subject to sanctions. Other conundrums are also identified, including that an extremely well functioning legal system may have most of its findings of liability be erroneous and a purely random legal system can have a perfect score (all findings of liability being correct).

In sum, conventional burden of proof formulations are deeply problematic in ways that have gone unnoticed, and these defects are in addition to the central one already noted: that their determinants have almost nothing to do with what evidence threshold maximizes social welfare. These conclusions motivate some inquiry into whether conventional proof burdens—in their canonical, idealized form, as considered throughout this Article—completely determine factfinders' actual behavior or whether, in practice, adjudicators are influenced to some nontrivial degree by attention to the consequences of their decisions. Also, some highly preliminary remarks are offered on alternative formulations of the burden of proof that are more attentive to social welfare.

The remainder of the Article returns to the central question of what evidence threshold is optimal and introduces a range of further considerations. One set pertains to other features of the legal system. First, this portion explores how to set the evidence threshold optimally when it is also possible to adjust the level of enforcement effort: a reduced threshold and heightened enforcement are substitute means of achieving a given increment to deterrence. A reduced threshold has the advantage of avoiding the additional expense involved with stronger enforcement but the disadvantage of faring relatively worse with regard to chilling effects. The relationship between the optimal evidence threshold and the level of sanctions is also addressed, and some similar interactions are identified. Attention is also devoted to how the direct costs of sanctions, such as with imprisonment, influence the optimal threshold. One might have thought a heightened threshold preferable because this would reduce the imposition of costly sanctions, but again this effect is only half of the story. The higher threshold, by reducing both deterrence and chilling, increases the number of cases that flow into the system, which tends to increase the rate of imposition of costly sanctions. Similar analysis applies to costs that might be associated per se with sanctioning the innocent: a higher evidence threshold makes that outcome less likely in a given case but tends to increase the number of cases involving innocents, so the net impact could be in either direction. The interplay between the accuracy of the legal system and how optimally to set the evidence threshold is also examined.

Finally, the Article considers similarities and differences across alternative enforcement environments. Monitoring (such as by the posting of police officers and use of patrols or private security guards) has effects on behavior that are similar to those of auditing (including various forms of inspection used to enforce myriad regulations), but the two methods differ in enforcement costs. Specifically, both deterrence and chilling, by reducing the level of activity, reduce the number of audits that need to be performed, even for a given audit rate. Enforcement by investigation, such as when a harmful act has clearly been committed (murder, auto theft) and officials attempt to identify the perpetrator, is notably different. The key point is that deterrence is more valuable: when fewer harmful acts are committed, fewer investigations are triggered, which both saves enforcement costs and also reduces the number of instances in which benign behavior may mistakenly be caught in the dragnet. An implication of the latter is that it seems entirely conceivable that demanding stronger evidence will increase the frequency with which innocent individuals are mistakenly sanctioned. Here the channel is that a tougher evidence threshold dilutes deterrence, which provokes more investigations and thereby creates more opportunities for mistakes to occur. This effect need not dominate the reduction due to the reduced likelihood that any given individual before a tribunal will be sanctioned, but it could.

A much different mode of enforcement involves the regulation of future conduct, such as when authorities decide whether to grant a zoning variance, license an individual or a facility, approve a drug, or allow a merger. In these settings, the determination of the optimal evidence threshold is more like traditional cost-benefit analysis, and it is perhaps most familiar from how one might decide whether to administer medical treatment following a diagnostic test. Unlike in the rest of the Article, the relevant factors differ much less from the determinants of the preponderance rule. Nevertheless, the differences that remain—namely, that the preponderance rule and other likelihood-based rules do not depend on harm and benefit—are of great consequence, making the traditional conceptions very poor proxies for guiding system design. Again, optimal evidence thresholds could be much stricter or notably weaker than any of these conventional rules dictates. Interestingly, in many of these contexts, legal systems often seem to proceed more in accord with what is socially optimal, eschewing customary burden of proof formulations.

The incentives of enforcers—public prosecutors, police, and agencies, as well as private plaintiffs—may influence the analysis as well. Such individuals and entities often seek to advance their own interests rather than the social good. This fact may be relevant in setting the evidence threshold both because enforcers' incentives influence the mix of cases that will come before tribunals (which mix affects what is optimal) and because the height of the threshold

may have feedback effects that influence, for better or worse, which cases are pursued.

This Article constitutes a conceptual investigation of the burden of proof, specifically analyzing how to set the evidence threshold so as to maximize social welfare. Although the central question is normative, much of the analysis is positive, concerning the effects of changing the evidence threshold in a variety of settings. Formulations are offered for the optimal threshold, indicating what factors are relevant and how they are to be combined. Sharp contrasts are drawn between the present analysis and conventional thinking about the burden of proof.

The investigation is, nevertheless, preliminary. It turns out that many of the identified effects, contrasts, and conclusions have received little prior attention. And a number are counterintuitive, in some cases rendering traditional understandings incoherent—in addition to indicating their great divergence from the advancement of social welfare. Introducing additional system features—enforcement effort, the magnitude of sanctions, and the degree of accuracy—and alternative enforcement contexts indicates numerous ways in which the results would change. A first cut at so complex and multifaceted a problem is inevitably incomplete, and other factors and settings remain to be examined. Perhaps most worrisome, many of the key determinants of the optimal evidence threshold seem quite difficult to quantify in any particular situation, which is all the more troubling given the extent to which they probably vary with the circumstances.

In the end, however, the only way forward is to fill in the many remaining blanks in the analysis, refine our thinking, and do our best to estimate pertinent empirical parameters. If we wish the legal system to better serve its function of controlling harmful activity while also taking into account costs due to the mistaken imposition of sanctions—that is, to maximize social welfare—there is no real alternative. Conventional conceptions of the burden of proof do not constitute even weak proxies for welfare because they depend almost entirely on factors different from those that determine the optimal evidence threshold. Moreover, they are vastly more complex to apply than is appreciated and raise a number of conundrums that cannot be avoided, except by refocusing our inquiry in a welfare-maximizing direction. In short, analysis of the burden of proof needs to start from scratch, and it is hoped that the present investigation will help instigate that effort.