THE YALE LAW JOURNAL

EDWARD K. CHENG

Reconceptualizing the Burden of Proof

ABSTRACT. The preponderance standard is conventionally described as an absolute probability threshold of 0.5. This Essay argues that this absolute characterization of the burden of proof is wrong. Rather than focusing on an absolute threshold, the Essay reconceptualizes the preponderance standard as a probability ratio and shows how doing so eliminates many of the classical problems associated with probabilistic theories of evidence. Using probability ratios eliminates the so-called Conjunction Paradox, and developing the ratio tests under a Bayesian perspective further explains the Blue Bus problem and other puzzles surrounding statistical evidence. By harmonizing probabilistic theories of proof with recent critiques advocating for abductive models (inference to the best explanation), the Essay bridges a contentious rift in current evidence scholarship.

ESSAY CONTENTS

INTRODUCTION

As every first-year law student knows, the civil preponderance-of-the-evidence standard requires that a plaintiff establish the probability of her claim to greater than 0.5.[1] By comparison, the criminal beyond-a-reasonable-doubt standard is akin to a probability greater than 0.9 or 0.95.[2] Perhaps, as most courts have ruled, the prosecution is not allowed to quantify "reasonable doubt,"[3] but that is only an odd quirk of the math-phobic legal system. We all know what is really going on with burdens of proof, especially with respect to 0.5.

But are these time-honored quantification moves actually correct? Is preponderance really $p > 0.5$ and beyond a reasonable doubt really $p > 0.95$? One need not dig too deeply to find immediate problems. Take, for example, the so-called Conjunction Paradox, which has long bedeviled legal scholars attempting to place the process of proof on probabilistic foundations.[4] Assume that a court is faced with a conventional negligence claim in which the plaintiff seeks to prove that: (*A*) the defendant was driving negligently; (*B*) the defendant's negligence caused him to crash into the plaintiff; and (*C*) the plaintiff suffered a soft-tissue neck injury as a result. Assume further that through the trial process, the plaintiff makes out each of these elements to a probability of 0.6. Should the plaintiff win? Each of the elements surely meets the preponderance standard; they all exceed 0.5. However, if all three elements are independent, their conjunction (*ABC*) has a probability of 0.6 * 0.6 * 0.6,

---

1. *E.g.*, Brown v. Bowen, 847 F.2d 342, 345 (7th Cir. 1988) (stating that under the preponderance standard, "the trier of fact rules for the plaintiff if it thinks the chance greater than 0.5 that the plaintiff is in the right"); Althen v. Sec'y of Dep't of Health & Human Servs., 58 Fed. Cl. 270, 283 (2003) ("[J]udges often express [preponderance of evidence] mathematically by saying the plaintiff must establish the facts necessary to her [or his] case by a probability greater than 0.5 or greater than 50%." (quoting DAN B. DOBBS, THE LAW OF TORTS 360 (2000))) (alterations in original).

2. *E.g.*, *Brown*, 847 F.2d at 345-46 (characterizing the beyond-a-reasonable-doubt standard as 0.9 or higher).

3. *See* Peter Tillers & Jonathan Gottfried, Case Comment, United States v. Copeland, *369 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim that Proof Beyond a Reasonable Doubt Is Unquantifiable?*, 5 LAW PROBABILITY & RISK 135, 135-36 (2006) (collecting cases).

4. *See, e.g.*, Michael S. Pardo, *Second-Order Proof Rules*, 61 FLA. L. REV. 1083, 1102 n.111 (2009) (classifying the Conjunction Paradox as an "analytical problem[] . . . for probabilistic rules"). *See generally* Ronald J. Allen & Sarah A. Jehl, *Burdens of Persuasion in Civil Cases: Algorithms v. Explanations*, 2003 MICH. ST. L. REV. 893 (suggesting that the Conjunction Paradox cannot be resolved by a theory or algorithm alone, but rather requires an additional substantive explanation).

or 0.216, suggesting that the plaintiff should lose. Even if the elements are not independent, their conjunction is always mathematically less than 0.6, so that with each additional element, the plaintiff finds it increasingly difficult to win.[5]

These types of problems present serious and fundamental impediments to scholars hoping to articulate a probabilistic theory of evidence.[6] They arguably even inhibit attempts to use probability and statistics to improve legal decisionmaking. After all, as it currently stands, the mathematics do not adequately model the legal system in operation. Along these lines, Ron Allen and Mike Pardo, among others, have argued that the legal system does not engage in this type of probabilistic reasoning at all, but instead proceeds through abductive reasoning, also known as inference to the best explanation.[7] Consistent with the story model of jury decisionmaking made famous by Nancy Pennington and Reid Hastie,[8] Allen and Pardo suggest that jurors choose the best explanation for the evidence with which they are presented. They do not accumulate evidence through conventional probability models.

But how could this state of affairs possibly be? On the one hand, probabilistic models of inference have been incredibly successful in science, leading to dramatic insights and findings into the way the world works. On the other hand, inference to the best explanation is compelling and intuitively correct to any lawyer. From law school on, lawyers learn that presenting a sagaciously chosen core theory (in appellate argument) or telling a compelling story (in trial argument) is critical to legal success.[9] Is legal factfinding simply different from scientific factfinding?

---

5. When the elements are not independent, their joint probability is the product of the probability of the first element and the *conditional* probability of the second element given that the first element is true. But assuming that the first two elements are not perfectly correlated, both probabilities must be less than 1, and the product is then necessarily less than the original separate probabilities. Mathematically, since $P(A|B) < 1$ and $P(B) < 1$, then $P(AB) = P(A \mid B) * P(B) < P(B)$.

6. *See generally* Ronald J. Allen, *A Reconceptualization of Civil Trials*, 66 B.U. L. REV. 401, 404-15 (1986) (reviewing the critiques of probabilistic models of evidence); Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. REV. 439, 450-67 (1986) (summarizing the criticisms against probabilistic models and the responses to them).

7. Michael S. Pardo & Ronald J. Allen, *Juridical Proof and the Best Explanation*, 27 LAW & PHIL. 223 (2008) (describing inference to the best explanation and its advantages over probabilistic models).

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

9. *See, e.g.*, THOMAS A. MAUET, TRIAL TECHNIQUES 24 (8th ed. 2010) ("A theory of the case is a clear, simple story of 'what really happened' from your point of view. . . . Trials are in large part a contest to see which party's version of 'what really happened' the jury will accept as

In this Essay, I argue that the answer to this question is in fact no. The use of probabilistic tools and the story model are not as antithetical as they may first appear. Indeed, the problem is neither in the use of probabilistic reasoning, nor in the use of a story model, but rather in the legal system's casual recharacterization of the burden of proof into $p > 0.5$ and $p > 0.95$. Indeed, once we recognize that mistake, we can construct a model of legal decisionmaking that is both compatible with the story model and potentially based on probabilities. As proponents of the story model have long argued, the legal system does not ask decisionmakers to determine whether litigants have established their cases to a particular level of certainty. Instead, decisionmakers compare the stories or theories put forward by the parties, and determine which story is more compelling in light of the evidence. However, far from calling into doubt the viability of probabilistic theories of evidence, this comparative procedure is found at the heart of standard methods of hypothesis testing in statistics. To make the two harmonize, evidence scholars need only let go of their love for $p > 0.5$.

The discussion proceeds as follows. In Part I, the Essay reconceptualizes the preponderance standard. It proposes viewing preponderance not as an absolute probability, such as 0.5, but rather as a ratio test that compares the probability of the narratives offered by the plaintiff and defendant.[10] With a probability ratio test in hand, later Sections show how the 0.5 standard came to be—essentially as an oversimplification—and how the ratio test avoids the Conjunction Paradox.

Part II pushes further on the reconceptualized preponderance standard by employing a Bayesian perspective. This Bayesian perspective offers a method of incorporating evidence into the decisionmaking process, and it provides an

---

more probably true."); *id.* at 64 ("Effective opening statements, like so much of trial work, are usually based on good storytelling.").

10. Probability ratios, of course, are not new to the evidence literature; indeed, they are a mainstay of discussions of probabilistic evidentiary models. *See, e.g.*, Richard D. Friedman, *A Presumption of Innocence, Not of Even Odds*, 52 STAN. L. REV. 873 (2000). The difference, however, is that conventional legal treatments focus on the likelihood ratio between the plaintiff's story being true and the plaintiff's story being false (as opposed to the defendant's story being true). *See, e.g.*, Louis Kaplow, *Burden of Proof*, 121 YALE L.J. 738, 773-74 (2012) (discussing a ratio in which acts are broken down into either "harmful acts" or "benign acts"); Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates*, 67 U. COLO. L. REV. 859, 869 (1996) (discussing a ratio involving a DNA match versus no match). As elaborated in Section I.A, this subtle shift is critical. Indeed, as Ron Allen insightfully understood at a relatively early point in these debates, it is this comparison of "the probability of the plaintiff's elements to that of their negation" that is the key problem behind current probabilistic theories of evidence. Allen, *supra* note 6, at 425.

explanation for the Blue Bus problem famous in statistical proof circles. Part III critiques the reconceptualized preponderance standard on normative grounds, departing from the otherwise explanatory goals of the Essay. As it turns out, the preponderance test as implemented by the legal system neglects base rates, which may explain the base rate problem's frustrating persistence. Part IV tentatively extends the ideas from the preceding Parts into the criminal context, and a Conclusion follows.

I. COMPARISONS, NOT ABSOLUTES

Conventional legal thinking equates the preponderance standard in civil litigation with a requirement that the plaintiff prove her case to a probability greater than 0.5. This Part argues that this characterization is wrong. Because the adversarial structure of legal trials promotes jury comparisons of the parties' claims, preponderance is not an absolute probability. Rather, the preponderance standard is better characterized as a probability ratio, in which the probability of the plaintiff's story of the case is compared with the defendant's story of the case. Indeed, while one can technically derive the $p > 0.5$ standard from the ratio, it involves assumptions sharply at odds with current legal practice.

Looking at the statistical world, we immediately see that characterizing any decision rule as a 0.5 probability threshold is odd. Statisticians rarely attempt to prove the truth of a proposition or hypothesis by using its absolute probability. Instead, hypothesis testing is usually comparative.[11] There is a null hypothesis and an alternative hypothesis, and one is rejected in favor of the other depending on the evidence observed and the consistency of that evidence with the two hypotheses.[12]

If one were to model the preponderance standard statistically, the natural move would therefore *not* be a 0.5 probability threshold. Rather, following

---

11. *E.g.*, E.L. LEHMANN & JOSEPH P. ROMANO, TESTING STATISTICAL HYPOTHESES 56-57 (3d ed. 2005) (defining the hypothesis-testing problem as being between a class *H* and an alternative class *K*).

12. Readers familiar with classical hypothesis testing may immediately note that classical hypothesis testing strongly favors the null hypothesis, a preference at odds with the usual practice in civil litigation. This preference is not always the case. For example, as described below, the null hypothesis can be given no specific preference. These complications, however, should not detract from the basic point, which is that hypothesis testing is typically comparative, just like in the story model.

standard decision theory, it might look something like this[13]: We start with two competing views of the world—for example, that the average height of an adult male is either 5'9" or 6'2" or, more generally for our purposes, that either the defendant's story (H_Δ) or the plaintiff's story (H_π) is true. Depending on the actual state of the world, concluding that the world is either H_Δ or H_π has error costs, as depicted in Figure 1. If the conclusion matches the truth, then obviously there is no error cost. However, if we conclude H_π but the world is actually H_Δ, we incur the costs of c_1. For the reverse, we incur the costs of c_2.[14]

Figure 1.

ERROR COSTS OF DECISIONS

CONCLUSION

		H_Δ	H_π
TRUTH	H_Δ	0	c_1
	H_π	c_2	0

Our goal is to construct a decision rule that minimizes our expected error costs. Let q_Δ and q_π represent the probabilities at which the states of the world H_Δ and H_π occur, respectively. We can then make some useful calculations about expected costs. For example, whenever we choose H_Δ, our expected costs will be:

$$0 \star q_\Delta + c_2 \star q_\pi$$

Similarly, if we choose H_π:

$$c_1 \star q_\Delta + 0 \star q_\pi$$

To minimize the expected costs, we will choose H_π if its expected costs are lower than those of H_Δ, or in other words if:

$$c_2 \star q_\pi > c_1 \star q_\Delta$$

---

13. The explanation derives from a useful presentation by Jonathan Rougier. *See* Jonathan Rougier, *Statistics 2: Bayesian Hypothesis Testing*, U. OF BRISTOL SCH. OF MATHEMATICS (Dec. 2008), http://www.maths.bris.ac.uk/~mazjcr/stats2/HOBayes2.pdf.

14. In labeling the error costs with c_1 and c_2, I have simply followed the Type I (false positive)/Type II (false negative) distinction commonly used, using the defendant's story as the baseline.

or equivalently, if:

Equation (1).

$$\frac{q_\pi}{q_\Delta} > \frac{c_1}{c_2}$$

Now, how do we determine these values? To estimate the probabilities q_π and q_Δ, we use all of the evidence: $P(H_\pi \mid E)$ and $P(H_\Delta \mid E)$, respectively, where E represents all of the available (or presented) evidence. At the same time, in a civil trial, the legal system expresses no preference between finding erroneously for the plaintiff (false positives) and finding erroneously for the defendant (false negatives). The costs c_1 and c_2 are thus equal, resulting in the decision rule that plaintiff wins if and only if:

Equation (2).

$$\frac{P(H_\pi \mid E)}{P(H_\Delta \mid E)} > 1$$

A. *Explaining the 0.5 Standard*

If the preponderance standard is actually a probability ratio, then where does the 0.5 number come from? After all, 0.5 seems awfully intuitive, which is arguably why it is so popular among lawyers. As it turns out, 0.5 arises from an error in assumptions.

Assume that the defendant's theory of the case is merely that the "plaintiff's theory is false." In the set of all possible stories of what happened, the plaintiff would have one story (justifying recovery) and the defendant would have all others simultaneously. Under these conditions, the defendant's set of stories is the complement of the plaintiff's set; in other words, $H_\Delta = H_\pi^c$. Then, from the axioms of probability,

$$P(H_\Delta \mid E) = P(H_\pi^c \mid E) = 1 - P(H_\pi \mid E)$$

And the decision rule, Equation (2), for when to accept H_π becomes:

$$\frac{P(H_{\pi} \mid E)}{1 - P(H_{\pi} \mid E)} > 1$$

which is equivalent to:

$$P(H_{\pi} \mid E) > 0.5$$

Thus, if we make the assumption that the plaintiff presents a single story that demonstrates the defendant's liability, *and* that the defendant may merely poke holes in the plaintiff's story, then we have the 0.5 standard. But this is neither how trials are structured, nor how juries operate, nor an intuitively attractive way to determine the truth. The defendant, particularly in a civil case, may not simply be a contrarian. The jury expects the defendant to present an alternative view of the evidence, and so like the plaintiff, the defendant too must present an explanation of what happened. To the extent that civil trials are about factfinding or truth, it will not do for the defendant's theory to be "not plaintiff's story." The defendant may offer multiple possible alternatives, but each of these alternatives will be judged separately, not simultaneously.[15]

There are of course instances in which 0.5 is a proper translation of the preponderance standard, but these are mere coincidences. For example, suppose that in a car accident case all facts are conceded except for the issue of whether the defendant was speeding. The plaintiff's story is that the defendant was speeding; the defendant's story is the complement, that he was not speeding. When the inquiry is this simple, and when the hypothesis is stated so that the plaintiff and the defendant take mutually exclusive and jointly exhaustive stances, then the 0.5 standard works. However, when the inquiry is more complex (for example, when the plaintiff must prove multiple elements or seeks relief on multiple claims), the defendant's alternative story is not a simple negation of the plaintiff's story, and the 0.5 standard is nothing but a source of confusion. The reconceptualized standard, by contrast, remains steadfast throughout these complications.

---

15. Allen, *supra* note 6, at 425 (proposing that civil trials should be reconceptualized "as comparing the probability of the fully specified case of the plaintiff to the probability of the equally well specified case of the defendant"). Some have argued that forcing the defendant to present "a competing version of the truth" diverges from current law. Kevin M. Clermont, *Death of Paradox: The Killer Logic Beneath the Standards of Proof*, 88 NOTRE DAME L. REV. (forthcoming 2013) (manuscript at 55-56) (on file with author). A natural response, however, is that if the defendant fails to provide a narrative, the jury will simply substitute the best narrative it can construct in favor of the defendant.

B. *Resolving the Conjunction Paradox*

Setting up the preponderance standard as a probability ratio yields immediate returns, as it solves the Conjunction Paradox, one of the serious challenges leveled against conventional probabilistic models of evidence. As previously discussed in the Introduction, even if the probability of each element of a claim is established to some level p, the probability of their conjunction (their simultaneous occurrence) will be no greater than p, and is often considerably lower; if the three elements of a tort are each established to the preponderance standard, say $p = 0.6$, and the three elements are statistically independent, the probability of all three occurring together is only $p_{all} = 0.6 * 0.6 * 0.6 = 0.216$, which is less than the 0.5 threshold. Claims with multiple elements are therefore potentially harder (often much harder) to establish.

To illustrate, consider a simplified version of the car accident case offered in the Introduction. The plaintiff claims that the defendant was speeding and that the crash caused her neck injury. The defendant responds that he was not speeding and that the plaintiff's neck condition was preexisting. The fact that the plaintiff has a neck problem is not disputed. In order to win, standard tort doctrine states that the plaintiff must establish that the defendant was speeding (breach of the duty of care), and that the breach caused the neck injury (causation). Although probably not strictly true empirically, let's presume both elements are statistically independent to simplify the calculations. In order to establish the entire case to the 0.5 preponderance standard, the plaintiff would have to establish each of the two elements to $p = 0.71$, since $0.71 * 0.71 \approx 0.5$.[16] It seems odd, however, that merely disputing another element of the tort not only creates a burden on the plaintiff regarding that element, but also raises the standard by which the plaintiff must prove both elements at issue.

The Conjunction Paradox, however, does not arise under the reconceptualized standard. To be succinct, let's call the speeding issue S and the causation issue C. Suppose that the plaintiff meets the burden of proof on both standards, so that we have probability ratios:

---

16. To be precise, the plaintiff might prove one element to $p > 0.71$ and the other to $p < 0.71$ so long as their product is still greater than 0.5, but the point is the same.

$$\frac{P(S\,|\,E)}{P(\overline{S}\,|\,E)} > 1$$

$$\frac{P(C\,|\,E)}{P(\overline{C}\,|\,E)} > 1$$

where E denotes the evidence in the case and the bar on top of the letter denotes negation, so that $\overline{S}$ means "not speeding," and $\overline{C}$ means "no causation." Note that as just argued in Section I.A, the defendant's story is not *necessarily* the negation of the plaintiff's, but in this case, the defendant was either speeding or not, and the impact either caused the injury or not.

To examine the case as a whole, we then look at conjunctions of those elements. The plaintiff's case is straightforwardly constructed. It is $S \cap C$, i.e., the defendant was speeding, and the impact caused the neck injury. The defendant's case is more complicated, because to win the defendant need only negate one of the elements. In other words, the defendant can win if (1) the defendant was speeding, but the plaintiff's neck injury was preexisting ($S \cap \overline{C}$); (2) the defendant was not speeding, but the impact did cause the injury ($\overline{S} \cap C$); or (3) the defendant was not speeding, and the plaintiff's condition was preexisting ($\overline{S} \cap \overline{C}$). However, the legal system wants the jury to arrive at some narrative of the truth. Any decision rule must thus consider these possibilities separately. As we see below, however, assuming that the plaintiff meets the preponderance standard (Equation (2)) on each element, the plaintiff will meet the standard for the entire case.

Scenario 1:

$$\frac{P(S\cap C\,|\,E)}{P(S\cap \overline{C}\,|\,E)} = \frac{P(S\,|\,E)P(C\,|\,E)}{P(S\,|\,E)P(\overline{C}\,|\,E)} = \frac{P(C\,|\,E)}{P(\overline{C}\,|\,E)} > 1$$

Scenario 2:

$$\frac{P(S\cap C\,|\,E)}{P(\overline{S}\cap C\,|\,E)} = \frac{P(S\,|\,E)P(C\,|\,E)}{P(\overline{S}\,|\,E)P(C\,|\,E)} = \frac{P(S\,|\,E)}{P(\overline{S}\,|\,E)} > 1$$

Scenario 3:

$$\frac{P(S\cap C\,|\,E)}{P(\overline{S}\cap \overline{C}\,|\,E)} = \frac{P(S\,|\,E)P(C\,|\,E)}{P(\overline{S}\,|\,E)P(\overline{C}\,|\,E)} > 1$$

It is worth noting that the intuition behind these calculations shares a kinship with the best attempts in the legal literature to address the Conjunction Paradox. For example, as Charles Nesson has noted, if the plaintiff can establish independent elements A and B each to 0.6, then while their conjunction falls below 0.5 (specifically, $P(AB) = P(A)P(B) = 0.36$), that state of affairs has a higher probability than any other combination.[17] The problem is that under Nesson's "solution," the decision rule is still nominally the 0.5 threshold, and the conjunction of elements that meet the 0.5 threshold can (and often do) result in probabilities below the 0.5 threshold. Under the reconceptualized burden of proof presented here, the same test (the probability-ratio-greater-than-1 test) allows movement seamlessly from case elements to the case as a whole.

C. *Story Definition*

One complication with the reconceptualized preponderance standard is the problem of story breadth. For example, in the previous example in Section I.B, the defendant's story was the more inclusive "not speeding" as opposed to the more specific "I was going 39 mph in a 40 mph zone." At the same time, the defendant's story was not allowed to be simply "not liable" or "not the plaintiff's story." The permissible breadth of a story is crucial, since the probability that the defendant was not speeding is obviously much higher than the probability that the defendant was going exactly 39 mph—the former includes not only 39 mph, but also 38, 37, and so on. So just how granular may a "story" be? Or put another way, why are some kinds of aggregation implicitly permitted, while other kinds are forbidden?

Arguably, the law answers this question directly through its structure. Aggregation may occur *within* a given legal element but not *across* legal elements.[18] Thus, on the issue of breach, the defendant's story may be as broad as "I was driving carefully," because that story is contained within the breadth of the breach element. A valid story, however, may not vaguely assert that either the defendant was not speeding or the plaintiff's injury was preexisting. Instead, as in the example above, the jury must consider each of the combinations separately: (1) speeding, preexisting; (2) not speeding, preexisting; and (3) not speeding, not preexisting. Similarly, a plaintiff may

---

17. $P(\overline{A}B) = 0.24$, $P(A\overline{B}) = 0.24$, and $P(\overline{A}\overline{B}) = 0.16$. *See* Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357, 1389-90 (1985) (developing this conjunction argument).

18. *See generally* Ronald J. Allen, *Factual Ambiguity and a Theory of Evidence*, 88 NW. U. L. REV. 604, 609-12 (1994) (discussing the aggregation problem for relative plausibility theory).

not bring both contractual and tort claims and then argue that regardless of the outcome on each individual claim, she should recover something because their disjunction (logically, contract OR tort) is greater than 0.5.[19]

II. BAYESIAN HYPOTHESIS TESTING

Thus far, the setup has been (hopefully) straightforward. Following the lead of the trial-practice literature and of proponents of inference to the best explanation,[20] we can reconceptualize the preponderance standard as a probability ratio test that avoids the Conjunction Paradox. That result is entirely general. As long as preponderance is a ratio test, the Conjunction Paradox disappears.

To extract additional insights about legal proof, this Part now looks at the reconceptualized standard through a Bayesian lens—specifically, using Bayesian hypothesis testing as a model.[21] As we will see, viewed in this light, the reconceptualized standard also helps explain a number of other puzzles surrounding statistical proof in the legal system.

Recall again the reconceptualized preponderance standard found in Equation (2). The plaintiff wins if and only if:

$$\frac{P(H_\pi \mid E)}{P(H_\Delta \mid E)} > 1$$

This form of the standard coheres well with our intuitions, but it offers little guidance on how to calculate the probabilities involved based on the available evidence. Here, Bayes' Rule can help.

We start with the prior odds, the starting ratio between the probability of one event (H_π) and the probability of another event (H_Δ). These odds are called "prior" because they are what we believe *prior* to observing the evidence. As one might imagine, prior probabilities or odds can be highly controversial. In statistical venues, critics argue that prior probabilities are overly subjective and

---

19. *See generally* Ariel Porat & Eric A. Posner, *Aggregation and Law*, 122 YALE L.J. 2 (2012) (discussing various methods of aggregating elements and claims).

20. *See supra* note 9.

21. The subsequent exposition is a relatively standard one found in texts or outlines on Bayesian statistics. *See, e.g.*, GEORGE CASELLA & ROGER L. BERGER, STATISTICAL INFERENCE 414 (2d ed. 2002); Rougier, *supra* note 13.

have no basis.[22] In legal venues, one might fear that they embody prejudices against certain types of parties. As we will see shortly, however, the use of prior odds ultimately does not raise such concerns in our analysis, so we can place them aside for the moment. We will denote prior probabilities of H_Δ and H_π as p_Δ and p_π. The prior odds are thus by definition:

$$\frac{p_\pi}{p_\Delta}$$

Bayes' Rule is the formula by which we "update" our prior beliefs by incorporating the evidence (E) that we observe. This update is accomplished through what is commonly called the Bayes Factor, which is nothing but a likelihood ratio.[23] It is the ratio between the probability of observing the evidence E given that H_π is true and the probability of observing the evidence E given that H_Δ is true. Thus, in mathematical terms, we have:

Posterior Odds = Bayes Factor ⋆ *Prior Odds*

$$\frac{P(H_\pi \mid E)}{P(H_\Delta \mid E)} = \frac{P(E \mid H_\pi)}{P(E \mid H_\Delta)} \star \frac{p_\pi}{p_\Delta}$$

In civil trials, the prior probabilities as a normative matter should arguably be equal.[24] As long as the plaintiff articulates a prima facie case and satisfies the

---

22. *E.g.*, Ronald Christensen, *Testing Fisher, Neyman, Pearson, and Bayes*, 59 AM. STATISTICIAN 121, 123 (2005) ("The absence of a clear source for the prior probabilities seems to be the primary objection to the Bayesian procedure.").

23. Here, I use the term "likelihood" for its precise statistical meaning, which is the probability of seeing the observed data under an assumed model. "Likelihood" should not, as in common parlance, be conflated with probability more generally.

24. *E.g.*, Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1514 (1999) (equating prior odds of 1:1 with an unbiased decisionmaker). Setting the prior odds ratio to 1 is potentially controversial. For example, Rich Friedman has criticized setting prior odds of 1:1, arguing that there is "no justification" that "simply because a proposition has been articulated, the proposition is exactly as likely to be true as false." Friedman, *supra* note 10, at 876; *see also* Kaplow, *supra* note 10 at 774 (noting that harmful acts and benign acts generally will not arise in court with "equal frequency"). Two responses, however, are in order. First, Friedman's article focuses largely on individual facts and thus instances in which prior odds of 1:1 may be more obviously questionable. *See* Friedman, *supra* note 10, at 877 (discussing dice or balls in an urn). When handling ultimate issues such as negligence, however, a normative legal position that starts the parties in equipoise has greater defensibility. Second, the goal here is to develop a probabilistic *model* of the legal factfinding process. Setting the prior odds at 1:1 may be wrongheaded as a matter of inference (indeed, Part III explores some of its costs), but that does not mean that courts do not do it.

burden of production,[25] the case starts with both parties in equipoise. The prior odds (p_π/p_Δ) are thus set at 1. The rule for deciding in favor of the plaintiff (H_π) therefore simplifies to:

$$\frac{\overparen{P(H_\pi \mid E)}}{P(H_\Delta \mid E)} = \frac{P(E \mid H_\pi)}{P(E \mid H_\Delta)} > 1$$

where the "hat" over the first probability ratio indicates that this is what the legal system uses to *estimate* the "true" value of:

Equation (3).[26]

$$\frac{P(H_\pi \mid E)}{P(H_\Delta \mid E)}$$

Let us take a moment to consider precisely what this new equation offers. Although the new test in Equation (3) may appear to merely "flip" the conditional probabilities in Equation (2), the difference is critical.[27] Under the original formulation, the reconceptualized preponderance standard asks the legal factfinder to compare, given the evidence, the probability of the plaintiff's story with the probability of the defendant's story. This setup makes intuitive sense, but it does not provide any guidance on how the factfinder might make such a comparison. The new test in Equation (3) suggests such a means (under a Bayesian framework). It is a kind of "likelihood ratio" test. The jury compares the likelihood of observing the evidence under the plaintiff's theory of the case with the probability of observing the evidence under the defendant's theory of the case. Again, whoever has the larger probability wins.

---

25. As it turns out, this assumption—that the burden of production is met—may be quite critical, for it excludes extreme cases in which both the plaintiff's and the defendant's narratives are ridiculous or unlikely. In those cases, juries (or perhaps the court on summary judgment) will find for the defendant regardless of the ratios.

26. Setting the prior odds to 1 for normative reasons necessarily means that the expression no longer equals $(P(H_\pi \mid E) \ / \ P(H_\Delta \mid E))$ in the strict mathematical sense. Rather, the expression is merely what the legal system uses as an estimate of the true value.

27. Indeed, flipping the conditional is often called the "transposition fallacy," and is an error to be avoided. In this case, however, setting the prior odds to 1 leads to the result.

A. *Resolving the Blue Bus and Gatecrasher Paradoxes*

The expression in Equation (3) gives us a possible explanation for the Blue Bus problem. Recall the facts of *Smith v. Rapid Transit, Inc.*,[28] as famously adapted by Laurence Tribe and Charles Nesson.[29] The plaintiff is driving along a two-lane undivided country road on a dark night when she is faced with the oncoming lights of a bus traveling along the median. To avoid an accident, the plaintiff swerves, causing her car to end up in a roadside ditch. Because of the emergency, the plaintiff is unable to observe anything except that the bus was blue. The plaintiff presents this testimony, along with evidence that the defendant, the Blue Bus Company, operates 80 percent of the blue buses in the town. The defense concedes both facts and presents no additional evidence. Is the plaintiff entitled to recover?

Facially, the plaintiff should win. The only evidence presented shows the probability that the defendant caused the accident is 0.8, which is clearly greater than 0.5. Yet, most people are uncomfortable with a verdict for the plaintiff, and more importantly, courts have generally rejected awarding damages to the plaintiff based on such "naked" statistical evidence.[30] But why? After all, isn't 0.8 greater than 0.5, plain and simple?

Once again, the 0.5 preponderance standard is the source of the apparent paradox. Rather than asking whether the probability that the defendant is responsible is greater than 0.5, we reconceptualize the preponderance standard as requiring a likelihood ratio greater than 1. So, in the Blue Bus case, we return again to Equation (3). Find for the plaintiff if and only if $P(E|H_\pi)/P(E|H_\Delta) > 1$, where H_π is the narrative that the Blue Bus Company owned the bus, and H_Δ is the narrative that another bus company did. E is the evidence observed, namely, that the plaintiff observed a blue bus and that the defendant operates 80 percent of the blue buses in town.

---

28. 58 N.E.2d 754 (Mass. 1945).

29. Nesson, *supra* note 17, at 1378-79; Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1340-41 & n.37, 1346-50 (1971).

30. *See, e.g.*, Howard v. Wal-Mart Stores, Inc., 160 F.3d 358, 360 (7th Cir. 1998) ("The plaintiff . . . asks for judgment on the basis of [statistical evidence] alone . . . ; he tenders no other evidence. If the defendant also puts in no evidence, should a jury be allowed to award judgment to the plaintiff? The law's answer is 'no.'"); Spencer v. Baxter Int'l, Inc., 163 F. Supp. 2d 74, 80 n.7 (D. Mass. 2001) ("The plaintiffs offer only 'naked statistical proof,' a type of evidence that the Massachusetts courts have found insufficient to support a jury verdict." (citation omitted)); Baker v. Bridgestone/Firestone Co., 966 F. Supp. 874, 876 (W.D. Mo. 1996) ("As a general rule, statistical evidence alone is insufficient to avoid a motion for directed verdict and necessarily a motion for a summary judgment.").

So the litigation hinges on the likelihood ratio $P(E|H_\pi)/P(E|H_\Delta)$. What exactly is this quantity in the Blue Bus case? The numerator is the probability of the evidence (that the plaintiff observes a blue bus and that the defendant operates 80 percent of the blue buses in town) given that the plaintiff was hit by one of the defendant's buses on the night in question. The denominator is the probability of the same evidence given that the plaintiff was driven off the road by a different company's bus on that dark night.

Recast in this way, the information that the defendant owns 80 percent of the blue buses in town, far from winning the case, borders on irrelevancy. It is hard to envision how the identity of the bus on the night of the accident, without more, gives us much, if any, information on the proportion of blue buses owned by the defendant.[31] If it gives us no information, then the likelihood ratio equals 1, and the plaintiff's evidence falls short of the decision rule, which requires a ratio greater than 1. Even if the identity of the wrongdoer happens to provide some information on the defendant's market share (after all, the mere fact that the defendant's blue bus was involved in an accident may raise the likelihood that the defendant operates a lot of the blue buses in town), the effect will be small, thus still explaining why we are uneasy with finding for the plaintiff despite the "clear" difference between 0.8 and 0.5.[32]

This analysis works on other naked-statistical-evidence hypotheticals. For example, in L. Jonathan Cohen's Gatecrasher Paradox, a rodeo is attended by 1,000 audience members, but the rodeo organizers sell only 499 admissions, meaning that 501 members of the audience are gatecrashers.[33] Assuming that payment was in cash and no receipt was given, can the rodeo organizers recover against a randomly selected audience member? Again, the raw probabilities under the traditional preponderance standard suggest yes, since there is a 0.501 probability that any randomly selected audience member is a gatecrasher. However, the reconceptualized preponderance standard suggests otherwise. The likelihood ratio is again $P(E|H_\pi)/P(E|H_\Delta)$. But whether the audience member is a lawful patron or a gatecrasher does not change the probability of observing the evidence presented. The likelihood ratio is therefore 1, and plaintiff fails to satisfy his burden of proof. Note that this result is the same if 750 or even 999 members of the audience are gatecrashers, in line with the

---

31. In an extreme case in which the defendant disclaimed owning any blue buses at all, the fact that the plaintiff was hit by the defendant's bus on the dark night might provide some sort of existence proof, but that is not the inquiry here.

32. This market share inference may also explain why some people who are uncomfortable with finding liability based on an 80 percent market share change their position when the numbers become 99 percent or 99.9 percent.

33. L. JONATHAN COHEN, THE PROBABLE AND THE PROVABLE 74-81 (1977).

slippery slope, even though we would empirically expect that at some point, most people would allow recovery against a randomly selected patron.

These analytic results offer an explanation for the hostility with which courts have historically treated naked statistical evidence, and why individualized "direct" proof matters so much for courts. Consider the effect on the Gatecrasher Paradox if a witness steps forward and testifies that she watched the defendant climb over the rodeo walls. Unlike with statistical evidence, the probability of observing such direct evidence changes (dramatically) under the plaintiff's and the defendant's theories of the case. The probability of an accusatory witness, given that the defendant was a gatecrasher, is clearly higher than the probability of a mistaken or perjurous accusatory witness, given that the defendant paid his fare. So the likelihood ratio quickly becomes greater than 1, and the plaintiff easily meets the burden of proof.

B. The Puzzle of Epidemiology

The Bayesian reconceptualized preponderance standard in Equation (3) explains a further puzzle related to the Blue Bus problem. Despite its stated abhorrence of statistical proof, the legal system has embraced epidemiological evidence — indeed, at times requiring epidemiology to prove general causation in toxic tort cases. This insistence upon epidemiology in toxic torts under the *Daubert* standard appears completely inconsistent with the desire for individualized evidence elsewhere.[34]

One can undoubtedly come up with explanations for the apparent shift. For example, the law seems most reluctant to use statistical proof when pinpointing the defendant's identity,[35] but there is no identity question in toxic tort cases, since the defendant is presumably already tagged on negligence or strict liability grounds. But there is no need to justify the distinction in terms of identity. The reconceptualized preponderance standard offers an alternative

---

34. *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993). One other area in which the legal system embraces statistical proof is in employment discrimination, but there the use of statistical evidence is arguably to prove a group-level issue — "pattern and practice" class action claims. In single-plaintiff cases, courts have rejected statistical proof "if the employer meets its burden of production and advances an individualized explanation for its conduct." Allan G. King, *"Gross Statistical Disparities" as Evidence of a Pattern and Practice of Discrimination: Statistical Versus Legal Significance*, 22 LAB. LAW. 271, 272 (2007).

35. *See* Tribe, *supra* note 29, at 1340-41. The Blue Bus and Gatecrasher Paradoxes involve questions of identity, as does perhaps the most famous case rejecting statistical proof, *People v. Collins*, 438 P.2d 33 (Cal. 1968), in which the court rejected statistical proof on various grounds — some statistical, some legal.

explanation why epidemiological evidence is relevant and readily received in legal decisionmaking, while Blue Bus-type evidence is not.

The explanation hinges again on the likelihood ratio in Equation (3):

$$\frac{P(E\,|\,H_\pi)}{P(E\,|\,H_\Delta)} > 1$$

In this case, the evidence (E) consists of the epidemiological studies showing an increase in disease risk for exposed populations. The element in dispute is general causation, so the plaintiff's story (H_π) is that the substance in question causes the disease, whereas the defendant's story (H_Δ) is that it does not.[36] Now consider what these values are. The probability of observing the evidence given the plaintiff's story is high: if the substance causes the disease, then one would expect epidemiological studies showing an increased risk from exposure. By contrast, the probability of observing such evidence given the defendant's story is comparatively low: If the substance is harmless, then seeing increased risks in well-conducted epidemiological studies is unlikely. The ratio is thus greater than 1, which satisfies the reconceptualized preponderance standard. Even more importantly, this likelihood ratio on the narrow issue of general causation then straightforwardly combines with likelihood ratios on other elements in the toxic tort case, as described in Section I.B.

III. OPTIMALITY

It is important to recognize that thus far, all I have shown is that reconceptualizing the preponderance standard away from the 0.5 threshold and toward a probability or likelihood ratio test is consistent with current practice and explains a number of puzzles that have historically accompanied attempts to apply formal probabilities to the trial process. This Part pursues the normative question: Is this decision rule optimal, or even desirable in some sense?

To start, we immediately face an important choice about the purpose of legal trials. If the purpose of legal trials is to maximize social welfare or some other economic criterion, then the focus on stories is almost certainly suboptimal, because it unnecessarily forces the factfinder to assess each story in

---

36. In this context, the defendant's negation would not appear to violate the rule requiring a specific story, since a substance is either capable of causing the disease or not. Even if one were to substitute a more specific story—e.g., that the plaintiff's disease was caused by some other known causal agent—the ultimate conclusion should still hold.

isolation, rather than make her best global guess.[37] However, assuming that the legal system's purpose is to discover a single, stable "truth" or narrative, then the current setup is correct. As a general matter, doing a hypothesis test based on likelihood ratios minimizes expected error costs when comparing hypotheses.

Nevertheless, the situation in the legal system is more complicated — and potentially more problematic. Recall that the legal system imposes a constraint on top of the standard Bayesian hypothesis testing setup. Normatively, it sets the prior odds ratio at 1 to start the plaintiff and the defendant in equipoise. On the one hand, this preordained prior odds ratio addresses the knotty problem of assigning priors, which has long been a major critique of Bayesian statistics.[38] On the other hand, setting the prior odds by decree carries undesirable effects. For example, it may cause legal actors to neglect base rates.

Perhaps the most well-known illustration of base rates is presented by Daniel Kahneman and Amos Tversky in their cab problem:

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
>
> (a) 85 percent of the cabs in the city are Green, and 15 percent are Blue.
> (b) A witness identified the cab as Blue.
>
> The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80 percent of the time and failed 20 percent of the time.
>
> What is the probability that the cab involved in the accident was Blue rather than Green?[39]

The answer to the question is 0.41, not 0.80. The proper probability calculation, which involves a straightforward application of Bayes' Rule, utilizes both the evidence observed (here, the witness) and the base rates (here,

---

37. *See* Porat & Posner, *supra* note 19, at 7 ("An act that is not clearly a strict liability tort and at the same time not clearly a negligence tort may nonetheless clearly be one or the other . . . and thus a wrongful act that should entitle the victim to a remedy.").

38. *See supra* note 22.

39. Amos Tversky & Daniel Kahneman, *Evidential Impact of Base Rates*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 153, 156-57 (Daniel Kahneman et al. eds., 1982).

that Blue cabs are relatively rare). Ignoring the base rates and looking only to the probabilities generated by the evidence is an error.

The Kahneman and Tversky example deals with absolute probabilities (precisely what this Essay has argued the legal system does not do), so having gained some intuition, let us consider an example involving res ipsa loquitur using probability ratios.[40] To get to the jury via res ipsa, the doctrine conventionally requires that the accident ordinarily not occur without negligence.[41] Translating that language into our probability framework, the standard becomes: given that the defendant took due care, the probability of the accident is low. In other words, $P(E|Nonneg) << 1$, where E is (as is often the case in res ipsa cases) nothing but the bare-bones evidence that the accident occurred. Implicitly, res ipsa also assumes that given negligent conduct, the probability of the accident occurring is higher, so that $P(E|Neg) > P(E|Nonneg)$. Based on these straightforward translations of the res ipsa rule, we have:

$$\frac{P(E|Neg)}{P(E|Nonneg)} = \frac{P(E|H_\pi)}{P(E|H_\Delta)} > 1$$

Since under Bayes' Rule,

$$\frac{P(H_\pi|E)}{P(H_\Delta|E)} = \frac{P(E|H_\pi)}{P(E|H_\Delta)} \star \frac{p_\pi}{p_\Delta}$$

then in any res ipsa case, we have:

$$\frac{P(H_\pi|E)}{P(H_\Delta|E)} > \frac{p_\pi}{p_\Delta}$$

Now, as previously discussed, as a normative matter, the legal system ordinarily sets the prior odds ratio (p_π/p_Δ) to 1. If so, that means that the plaintiff always wins (or more precisely, we expect that the jury will find for

---

40. This discussion of res ipsa was in part inspired by David Kaye, *Probability Theory Meets Res Ipsa Loquitur*, 77 MICH. L. REV. 1456 (1979), which looks at res ipsa through the perspective of probability theory.

41. See, for example, *Dupont v. Fred's Stores of Tenn., Inc.*, 652 F.3d 878, 882-84 (8th Cir. 2011), where a customer allegedly sustained a head injury when plastic bins fell off of a shelf; and *Byrne v. Boadle*, (1863) 159 Eng. Rep. 299 (Exch.) 300, a classic English case in which a barrel fell on the plaintiff. *See also* Restatement (Second) of Torts § 328D (1965).

the plaintiff). The problem is that this prior odds ratio is often empirically not 1. In a res ipsa case, p_π is the probability that people act negligently as a general matter, absent any evidence. This number is almost certainly small and, at a minimum, is much smaller than the probability that people act nonnegligently (i.e., (p_π/p_Δ) << 1).

So, if we adhere to a Bayesian model of proof in a res ipsa case, we should *not* automatically find in favor of the plaintiff because, depending on the base rates, the mere fact that an accident occurred does not necessarily make the plaintiff's tale of negligence more likely than the defendant's tale of innocence. However, because the legal system forces the prior odds to be 1, it only sees half the picture and finds liability far more often than it should.

The Bayesian reconceptualized preponderance standard thus exhibits base rate blindness.[42] But while this result is undesirable as a practical matter, it is quite illuminating as an explanatory matter. The noble and persistent efforts of evidence scholars urging courts to take greater account of base rates have often faced considerable resistance.[43] Typically, one might attribute the failure to the usual concerns: inertia, laziness, ignorance, etc. What the reconceptualized standard shows us, however, is that base rate blindness may be far more fundamental to legal factfinding. If fairness principles require courts to set the prior odds to 1, then the ultimate accuracy of the system will pay a price.

IV. AN EXTENSION TO CRIMINAL CASES

The most immediate and natural extension of the reconceptualized preponderance standard is to the criminal law, where the beyond-a-reasonable-doubt standard has often been (informally) quantified as 0.90 or 0.95. At first blush, however, the context seems considerably different. First, the tolerance for error is asymmetric. In the civil context, the legal system equally weighs error on either side. In the criminal context, courts strongly prefer avoiding wrongful convictions (false positives) even at the cost of acquitting the guilty (false negatives).[44] In fact, this preference is arguably not just a weighting of

---

42. *See* Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 808-09 (2001) (discussing base rate issues in the res ipsa context); Kaye, *supra* note 40, at 1473-74 (discussing the importance of prior probabilities in the res ipsa analysis).

43. *See* Jonathan J. Koehler, *When Do Courts Think Base Rate Statistics Are Relevant?*, 42 JURIMETRICS J. 373, 377 (2002) (describing cases in which courts have found base rates irrelevant).

44. *See, e.g.*, *In re* Winship, 397 U.S. 358, 371-72 (1970) (Harlan, J., concurring); Josh Bowers, *Physician, Heal Thyself: Discretion and the Problem of Excessive Prosecutorial Caseloads, A Response to Adam Gershowitz and Laura Killinger*, 106 NW. U. L. REV. 143, 152 n.34 (2011)

the two types of error. Rather, it is a commitment to keep the wrongful conviction rate below a prescribed level, which invokes an entirely different schema of hypothesis testing.[45] Second, the urgency of reconceptualizing the reasonable doubt standard is reduced because courts have flatly rejected quantification in criminal cases. Courts have regularly disfavored the quantification of reasonable doubt as potentially unconstitutional,[46] and with good reason, for, as we shall see, the criminal burden of proof arguably has nothing to do with getting the probabilities above 0.95.

Despite these differences, this Essay's basic insight remains valid. Rather than work with absolute probabilities and a quantified threshold (in this case, 0.95), the better way to model factfinding in criminal cases is again as a likelihood ratio.

While the civil burden of proof is well modeled as a species of Bayesian hypothesis testing, the criminal burden of proof, with its commitment to a predetermined, low false-positive rate, is arguably best modeled using classical hypothesis testing, which may be more familiar to those who took college statistics.[47] Classical hypothesis testing starts with a null hypothesis, in effect a

("[E]mbedded in the prosecutor's burden of proof beyond a reasonable doubt is a normative abhorrence for Type I errors (wrongful convictions) as compared to Type II errors (wrongful acquittals)."); Gregory M. Gilchrist, *Plea Bargains, Convictions and Legitimacy*, 48 AM. CRIM. L. REV. 143, 147 (2011) ("The presumption of innocence and the government's burden of proof beyond a reasonable doubt cause the evidence to be weighed in a manner that favors, by design, wrongful acquittals over wrongful convictions.").

45. *See* Boaz Sangero & Mordechai Halpert, *Proposal To Reverse the View of a Confession: From Key Evidence Requiring Corroboration to Corroboration for Key Evidence*, 44 U. MICH. J.L. REFORM 511, 544-46 (2011) (demonstrating how the reasonable doubt standard can be quantified in order to limit the number of wrongful convictions to a predetermined level).

46. *E.g.*, State v. Rizzo, 833 A.2d 363, 399 (Conn. 2003) (explaining that "it is improper for a trial court to attempt to explain the concept of reasonable doubt by metaphors or analogies that are quantified in nature" because it "dilute[s]" the "constitutional standard of proof beyond [a] reasonable doubt" (citing State v. DelVecchio, 464 A.2d 813, 818-19 (Conn. 1983))); Commonwealth v. Rosa, 661 N.E.2d 56, 63 (Mass. 1996) (admonishing judges "to avoid examples that have numeric or quantifiable implications"). *But see* Petrocelli v. Angelone, 248 F.3d 877, 888-89 (9th Cir. 2001) (finding that the trial judge's use of a "97 yard line" analogy did not violate due process when it was "an off-hand remark" and the judge gave correct instructions after voir dire and before deliberation). *See generally* Tillers & Gottfried, *supra* note 3, at 135-36 (cataloging cases rejecting quantifications of reasonable doubt).

47. This analogy between the criminal burden of proof and classical hypothesis testing is often found in the statistics literature for purposes of illustration. *See, e.g.*, Michael W. Trosset, Criminal Law and Statistical Hypothesis Testing (1997) (unpublished technical report), http://www.stat.purdue.edu/~yuzhu/stat514s2006/Supplement/law.ps. *But see* Tung Liu & Courtenay C. Stone, *Law and Statistical Disorder: Statistical Hypothesis Test Procedures and the Criminal Trial Analogy* (Dept. of Econ., Ball State Univ., Working Paper No. 200601, 2006), http://econfac.iweb.bsu.edu/research/workingpapers/bsuecwp200601r1liu.pdf

favored or status quo result from which we will depart only if we have considerable evidence to the contrary. In criminal justice, a finding of "not guilty" easily fills the role of a null hypothesis, with guilt being the natural alternative if we have sufficient evidence. In classical hypothesis testing, the false positive rate is set a priori to a fixed value, commonly labeled as α. For example, we might set α at 0.05—effectively accepting the chance of one wrongful conviction for every twenty innocents tried.

The goal then becomes finding a decision rule, given this α constraint, that maximizes the "power": the probability of finding guilt when the defendant is in fact guilty. But how to do that? Here, a time-honored statistical result, the Neyman-Pearson fundamental lemma, conveniently pertains.[48] Jerzy Neyman and Egon Pearson found that when one is testing a simple null hypothesis versus a simple alternative hypothesis—a single defense narrative of innocence versus a single prosecution narrative of guilt—the most powerful test has the following elegant form: reject the null hypothesis if the probability of the data under the alternative hypothesis divided by the probability of the data under the null is greater than some constant k. In other words, reject the null hypothesis if:

$$\frac{P(E \mid H_{alt})}{P(E \mid H_{null})} > k$$

Or in the criminal context, convict the defendant only if:

$$\frac{P(E \mid H_{\pi})}{P(E \mid H_{\Delta})} > k$$

We thus see a likelihood ratio test again. Notably, rather than the more straightforward 1, the threshold is now the somewhat mysterious k, but the form is identical. Determining k requires all kinds of assumptions about the underlying probability distributions, so precise calculation of k might be extremely difficult in any practical legal context. Perhaps in practice, the jury relies on its intuition in setting k. But the mathematical complications here are

---

(expressing concerns about using such analogies in attempting to teach statistics). *See generally* Michael J. Saks & Samantha L. Neufeld, *Convergent Evolution in Law and Science: The Structure of Decision-Making Under Uncertainty*, 10 LAW PROBABILITY & RISK 133 (2011) (showing the uncanny convergence in approach between statistical hypothesis testing and criminal adjudication).

48.   J. Neyman & E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, 231 PHIL. TRANSACTIONS ROYAL SOC'Y LONDON SERIES A 289 (1933).

beside the point. The key is that, once again, the way to model the criminal burden of proof is not with absolute probabilities. The better way is with a likelihood ratio, which coheres with the story model. In addition, the precedents decrying the quantification of reasonable doubt at 0.95, far from being a species of statistical Luddism, may be actually correct.

One objection to this formulation of criminal factfinding is that it seems to violate the presumption of innocence. In criminal cases, technically speaking, the defendant can put the prosecution to its proof—meaning that if the prosecution fails to provide sufficient evidence of guilt, the defendant is declared "not guilty," even if the defendant says nothing and provides no explanation. The likelihood ratio test would appear to require that the defendant offer a null hypothesis, in violation of this rule. This concern, however, is remedied by noting that if the defendant declines to provide the null, the jury can provide its own. More specifically, given the evidence, the jury constructs its own potential narratives as to what really happened and matches each of these narratives against the prosecution's theory in turn.

CONCLUSION

Equating the burden of proof to an absolute probability is a mistake. In the civil context, the preponderance-of-the-evidence standard is not an absolute probability greater than 0.5, but rather a probability ratio or likelihood ratio greater than 1. In the criminal context, the beyond-a-reasonable-doubt standard is similarly not an absolute probability of 0.95, but rather also a likelihood ratio test where the threshold is set through (contestable) notions of the acceptable false positive rate.

More importantly, reconceptualizing the burden of proof provides a way to harmonize probabilistic and story or explanatory theories of proof. At present, the evidence literature contains repeated and often harsh volleys between the probabilists and their critics. On the one hand, the critics are surely right. Probabilistic theories of evidence based on the 0.5 standard have profound and vexing problems that thus far have lacked satisfactory solutions. On the other hand, the explanatory theories of decisionmaking have their own problems. Among other things, how could statistics, a dominant modern field addressing the issue of inference, have little to contribute to proper decisionmaking in the legal system? Such a state of the world seems both odd and highly improbable.

Reconceptualizing the burden of proof as a likelihood ratio test shows that both sides are partly right. The probabilists are not misguided in attempting to use modern statistics to improve models of legal decisionmaking; their mistake is in the setup. Similarly, the critics are not incorrect in their critique; their mistake is to throw out the proverbial baby with the bathwater. The solution is

not to abandon one theory for the other. Rather, by using probability ratio tests, we harness both the comparative perspective advocated by explanatory theories of evidence and the rigor of probabilistic theories. Rather than a rift, we instead have a cohesive whole.

Much work remains to be done in giving a full account of the legal proof process, but hopefully this Essay has reset the parameters of the debate. Not only should courts and attorneys stop using the misleading 0.5 rule as a shorthand for the preponderance standard, but with probability ratios in tow, perhaps evidence scholars can get back to the grand task at hand rather than fighting amongst themselves.