A LIKELIHOOD STORY

THE THEORY OF LEGAL FACT-FINDING[*]

Sean P. Sullivan[†]

April 15, 2017

For over 50 years, courts and scholars have tried to explain fact-finding, and burdens of persuasion, in terms of the probability of the facts given the evidence. The exercise has not been a success. The problem is reliance on Bayesian posterior probabilities. Fact-finding is not about probability. It's about likelihood. The difference is more than semantic. Where Bayesian probability asks about the probability of the facts given the evidence, likelihood asks about the probability of the evidence given the facts. And where probability formalizes the concept of personal belief, likelihood formalizes the concept of weight-of-evidence alone. Using the statistical properties of likelihoods, I show that every burden of persuasion in use today can be reduced to the same rule of likelihood reasoning. This likelihood theory of fact-finding formalizes story-based descriptions of the cognitive process of fact-finders, and illustrates the smooth progression of story-based models to heightened burdens of persuasion. It also resolves many of the paradoxes that beset the Bayesian probability theory of fact-finding, and clarifies the nature of uncertainty and appropriate form of inference in many fact-finding applications.

[*] This paper has benefitted from the thoughtful comments of Ron Allen, Ed Cheng, Kevin Clermont, Ken Elzinga, Josh Fischman, Brandon Garrett, Mike Gilbert, Josh Goodman, Dan Greenfield, Jay Koehler, Greg Mitchel, John Pepper, Devesh Raval, Paul Rothstein, Bob Tovsky, and many workshop participants at the Federal Trade Commission, University of Iowa, and University of Virginia.
[†] University of Iowa College of Law. Contact the author at s-sullivan@uiowa.edu.

Table of Contents

"No lawsuit can be decided, rationally, without the application of the commonplace concept of burden of proof.... Nor can any legal system be praised for practicability if there exists vagueness, uncertainty or confusion as to the scope or extent of the burden, or if the language commonly employed to describe its scope or extent is not easily comprehensible to those whose duty it is to determine whether the burden has been sustained."—J.P. McBaine (1944)[1]

1 INTRODUCTION

I have two cups, each containing six marbles. One of them (the *white cup*) has five white marbles and one red; the other (the *red cup*) has five red marbles and one white. The cups are indistinguishable from the outside. Consider two simple puzzles. First, suppose you were to draw a marble at random from the white cup. What is the probability that you would draw a white marble? Second, suppose I choose a cup outside of your view, and then, without revealing which cup I chose, let you randomly draw a marble—which turns out to be white. Given this observation, what is the probability that I chose the white cup?

If you are like most people I talked to in writing this article, you can easily answer the first question: the probability of drawing a white marble from the white cup is 5/6 (about 83%). And if you are like most people I talked to, you would comfortably guess that I chose the white cup in the second question. Assigning a numeric probability to that choice is trickier; some people recall that there is a rule of probability for this calculation (Bayes' Theorem), but few know it intuitively.

And if you are like most people I talked to in writing this article, you've already guessed that one of these puzzles was a trick question. In the first puzzle, there was indeed a 5/6 probability of seeing a white marble on a future draw. But in the second puzzle, it doesn't really make sense to talk about the *probability* of my choice at all. The reason is obvious, once you think about it. For one thing, this not a future event with value still to be realized: I've already chosen the cup. That

[1] J.P. McBaine, *Burden of Proof: Degrees of Belief*, 32 CALIF. L. REV. 242, 242 (1944).

historic choice is now a fact in the world, not a random variable. The probability that I chose the white cup is—in a literal sense—either zero or one. More importantly, even if you did recall Bayes' Theorem, and tried to use it to calculate the probability that I had chosen the white cup, you would have found that you need to know something that I didn't tell you: the latent or *prior* probability that I would choose the white cup over the red cup. But I assigned no probability to this choice because it, too, was not random. I chose a cup by conscious act. There was nothing probabilistic about it, and thus no way for me to describe the probability of my choice short of simply answering the puzzle.

In retrospect, you might feel that the second puzzle was impossible to answer. But then ponder this: why, for the last 50 years or more, have courts and scholars struggled to describe fact-finding, and burdens of persuasion, in terms of the *probability* that the defendant committed a crime, acted negligently, or breached a contract? When the facts at issue are historic acts and conscious decisions—which do not reduce to meaningful probabilities—the challenge of trying to determine the probability of the facts is really no different than trying to assign probabilities to my choice of cup in the second puzzle.

So why the fixation on probability language in fact-finding? Part of the answer is certainly that—even today—we are still trying to develop a theory of fact-finding, and Bayesian probability analysis is only one of the more recent efforts to supply this framework. But while that explains the need for a theory of fact-finding, it does not explain the turn to probability for the answer. And even if probabilities could be calculated for the long-run or average truth of some set of facts, it is hard to see why it would matter in a legal setting. Responsibility for specific action is basic to our system of justice. We do not punish a murderer for having some average or probabilistic tendency toward violence, but for the specific act of raising and lowering the knife.

The logical disconnect in searching for historic *facts* by reference to the *probability* of their truth is itself troubling, but the problems of the probability approach go deeper still. Decades of attempts to build a Bayesian probability theory of fact-finding have succumbed to disturbing problems:[2] strange dependence on the way that elements of a cause of action are framed, unsettling sensitivity to prior probabilities, and an alarming inability to say what specific facts would be found in even the simplest of fact-finding exercises, to name a few.

These problems are well documented, and while some researchers still seek a Bayesian probability theory to fact-finding,[3] many have abandoned their search for a formal theory altogether, instead focusing on trying to better understanding fact-finders' cognitive process.[4] Exciting work on relative plausibility,[5] narrative coherency,[6] and story-based[7] models advances our understanding of how fact-finders think (iteratively constructing and comparing stories that explain and relate to the evidence), but does little to improve our understanding of the burdens of persuasion, and so is not a substitute for a formal theory of fact-finding. Nor is there reason to give up the search for a formal theory. The problem with efforts to explain fact-finding in terms of Bayesian probabilities is not that mathematics and probability theory are inherently irreconcilable with fact-finding.[8] The problem is simply that most fact-finding seems not to be about Bayesian probabilities at all. It's actually about *likelihoods*.

This article presents a likelihood theory of fact-finding. The difference from a Bayesian probability approach is more than semantic. At an intuitive level, Bayesian posterior probability represents what a fact-finder personally believes after seeing the evidence; likelihood formal-

---

[2] *See, e.g.*, Brian Leiter & Ronald J. Allen, *Naturalized Epistemology and the Law of Evidence*, 87 VA. L. REV. 1491, 1503–10 (2001) (discussing several such problems).
[3] *See, e.g.*, Edward K. Cheng, *Reconceptualizing the Burden of Proof*, 122 YALE L.J. 1254, 1256 (2013).
[4] *See generally* Michael S. Pardo, *The Nature and Purpose of Evidence Theory*, 66 VAND. L. REV. 547 (2013) (summarizing recent developments in evidence theory).
[5] *E.g.* Ronald J. Allen, *Factual Ambiguity and a Theory of Evidence*, 88 NW. U. L. REV. 604 (1994).
[6] *E.g.* Dan Simon, *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, 71 U. CHI. L. REV. 511 (2004).
[7] *E.g.* Nancy Pennington & Reid Hastie, *The Story Model for Juror Decision Making*, *in* INSIDE THE JUROR: THE PSYCHOLOGY OF JUROR DECISION MAKING 192 (Reid Hastie ed., 1993).
[8] *Cf.* Lawrence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971).

izes the more basic concept of what the evidence alone shows. Another way to see the difference between these concepts is in the cup-choice puzzle. Upon drawing a white marble, most people guess that I have chosen the white cup. They do not base this guess on Bayes' Theorem, but on the more basic logic that the white cup is most likely to have produced the evidence they've seen. That is, where Bayesian analysis concerns the probability of the facts given the evidence, likelihood analysis concerns the probability of the evidence given the facts.

I argue that most legal fact-finding (any search for historic facts and actions) is best approached and understood as likelihood reasoning. This is not the first paper to suggest the use of likelihood reasoning in fact-finding,[9] but it is the first paper to explain all current burdens of persuasion in terms of likelihood reasoning alone. To be clear, though, my argument is not that this or any formal theory of fact-finding should be forced on fact-finders in a trial setting. Nor do I argue that this theory captures every nuance of the fact-finding process; like any formal model, the theory is an abstraction with important limitations.

What I *do* emphatically argue is that a reasonable and internally coherent theory of fact-finding should guide our legal practice and procedure. My proposed likelihood theory serves this role. It improves upon the failed Bayesian approach to fact-finding, resolving the paradoxes of earlier theories. It simplifies legal analysis, describing every burden of persuasion in terms of the same rule of likelihood reasoning. It harmonizes the literature, formalizing empirical work on the cognitive processes of fact-finders. And it offers insights into the problem of subjective belief, clarifying the problem of fact-finder bias, in particular. The final promise of this likelihood theory is clearer legal thinking and a jurisprudence of fact-finding that at last reflects common sense and the realities of trial practice.

To solidify the stakes, consider one last puzzle, involving a claim of unjust enrichment, tried without a jury. Suppose that, before any evidence is presented, the judge tells you the defendant strikes him as the kind of person who would exploit a benefit; he puts an 80% probability on liability. But the plaintiff produces only weak evidence for her

---

[9] *E.g.* Louis Kaplow, *Burden of Proof*, 121 YALE L.J. 738 (2012).

claim, while the defendant puts forth strong evidence that she was not enriched, much less unjustly so. Everyone (even the judge) concedes that the weight of the evidence favors the defendant. Yet because his prior opinion was so strong, the judge nevertheless finds the defendant liable. This is a textbook example of how a Bayesian probability theory of fact-finding would work. But is this really how fact-finding works? Is it how it should work? If this process and its outcome strike you as unjust, unfair, and wrong, the principles of a likelihood theory of fact-finding may explain why.

2 THE LANDSCAPE OF UNCERTAINTY AND PERSUASION

To show how likelihood reasoning motivates a theory of fact-finding, I need to start with something the evidence literature has rarely given much attention: the difference between probabilities and likelihoods. This discussion is unavoidably formal, but intuition and application to the legal context are always at hand.

2.1 *Probability (Absolute Subjective Belief)*

Probability is an absolute description of the uncertainty in a system. It translates relative frequencies or beliefs into numbers between 0 and 1. For example, to say that a random marble drawn from the white cup has a 5/6 probability of being white means that if I were repeat this process an infinite number of times (draw a marble; note its color; replace it; repeat), 5/6 of these draws would be white, and 1/6 would be red. When probability statements depend on other factors, they can be expressed conditional on these factors. For example, the probability of drawing a white marble depends on which cup I chose:[10]

$$P(draw\ white\ marble \mid white\ cup\ chosen) = 5/6$$
$$P(draw\ white\ marble \mid red\ cup\ chosen) = 1/6$$

Decades of work on fact-finding have focused on trying to swap the order of variables and conditions in this type of probability statement.

[10] In typical notation, conditioning variables appear to the right of a vertical bar "|" in the probability statement. The notation $P(A|x)$ is read "probability of A given x."

The fact-finder is assumed to know (at least subjectively) the probability of seeing the evidence given the facts. And the objective is then to infer the reverse probability: the probability of the facts given the evidence. But, in general, these terms cannot be simply swapped:

$$P(\mathit{white\ cup\ chosen} \mid \mathit{draw\ white\ marble}) \neq 5/6$$

Instead, reversing the order of terms requires appeal to a fundamental law of probability: Bayes' Theorem.[11]

The most common use of Bayes' Theorem in modern fact-finding applications involves the relative probability of two events before and after evidence is observed:[12]

$$\underbrace{\frac{P(A|x)}{P(B|x)}}_{\substack{\text{posterior} \\ \text{dist.}}} = \underbrace{\frac{P(x|A)}{P(x|B)}}_{\substack{\text{likelihood} \\ \text{ratio}}} \times \underbrace{\frac{P(A)}{P(B)}}_{\substack{\text{prior} \\ \text{dist.}}}$$

The left-hand term in this equation (the posterior distribution) gives the relative probabilities of two events, A and B, after evidence x has been observed and considered. In the cup-choice puzzle, for example, the evidence is the white marble drawn from the unidentified cup, and the events are my possible choices of cup. The rightmost term in the equation (the prior distribution) describes the relative probabilities of these events *before* the evidence is observed. And the middle term (the likelihood ratio) conveys the relative consistency of the evidence with each of the possible events. So the relative probabilities after seeing the evidence, are equal to the relative probabilities before seeing the evidence, multiplied by the likelihood ratio of the events on the evidence (more on that later).

---

[11] Summary exposition of Bayes' Theorem can be found in any introductory text on probability and statistics. *E.g.* LEE J. BAIN & MAX ENGELHARDT, INTRODUCTION TO PROBABILITY & MATHEMATICAL STATISTICS 22–27 (2d ed. 1992). Though typically attributed to Thomas Bayes, the origins of the theorem are murky. *See generally* Stephen M. Stigler, *Who Discovered Bayes's Theorem*, 37 AM. STATISTICIAN 290 (1983).

[12] *See, e.g.*, Chris William Sanchirico, *Evidence: Theoretical Models*, *in* X ENCYCLOPEDIA OF LAW AND ECONOMICS *7–9 (2012).

The intuitive appeal of Bayes' Theorem in fact-finding applications is hard to miss: just replace A and B with *guilty* and *innocent* and you have the probability model of fact-finding that has dominated the evidence literature for more than 50 years.[13] But while much of Bayes' Theorem has been explained and re-explained in prior work,[14] one aspect of the model that always seems to get short shrift is the way that results depend on the source of the prior probability distribution.

Things are straightforward when an obvious, objective prior distribution is available. As an artificial example, suppose we were to repeat the cup-choice puzzle, but this time I based my choices on a coin flip: heads, I'd choose the white cup; tails, red. This randomization provides a clear prior distribution for my choice: there is a ½ prior probability that I will choose each cup. And in this context, upon randomly drawing a white marble, you could use Bayes' Theorem to calculate the newly informed relative probability that I had chosen the white cup:

$$\frac{P(\textit{white cup chosen} \mid \textit{draw white marble})}{P(\textit{red cup chosen} \mid \textit{draw white marble})} = \frac{5/6}{1/6} \times \frac{1/2}{1/2} = 5$$

The math is not as important as the interpretation—and in this case the interpretation is as a long-run ratio of frequencies. If we were to repeat this entire process an infinite number of times, then within the set of repetitions in which you drew a white marble, you should expect me to have chosen the white cup 5 times as often as the red cup. This is not so much a statement about my choice in any given repetition as it is a statement about the long-run average of my choices. In short, where objective long-run or average prior probabilities are available, the output of Bayes' Theorem is likewise a description of long-run or average probabilities.

But what happens when clear prior probabilities are not available? In the actual cup-choice puzzle, I do not flip a coin, but simply choose a cup. What prior probabilities describe this non-random choice? And

[13] *See* John Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065, 1083–91 (1968) (providing what appears to be the first work on this topic); *see also* Richard Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021 (1977).

[14] *See, e.g.*, Lempert, *supra* note 13, at 1022–25; Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1486–87 (1999).

in attempts to apply Bayes' Theorem in a legal context, what are the prior probabilities that a person would breach a contract, neglect a duty, or assault someone? Bayes' Theorem cannot function without these prior probabilities, but there is little hope of finding objective long-run or average probabilities in this context.[15]

When clear and objective prior probabilities are unavailable, Bayesian analysis falls back on subjective prior probabilities:[16] probabilities that merely reflect the individual beliefs of the user. The output of Bayes' Theorem is no longer a description of long-run or average probability in this case; it is more like a normative claim about beliefs. By combining a person's initial beliefs with the observed evidence, Bayes' Theorem describes what a fully rational person—who just happened to hold these prior beliefs—should now believe after seeing the evidence. But since prior beliefs will vary from person-to-person and context-to-context, the exact same evidence can produce starkly different posterior probabilities (beliefs) under this form of Bayesian probability analysis.

2.2 *Likelihood (Relative Weight of the Evidence)*

Though the terms are synonyms in informal English,[17] modern statistics draws a distinction between *probability* and *likelihood*. If Bayes' Theorem is concerned with the probability of an event given observed evidence, $P(A|x)$, then likelihood is concerned with the probability of observing the evidence given a hypothesized event, $P(x|A)$. Where

---

[15] *See* Ronald J. Allen & Alex Stein, *Evidence, Probability, and the Burden of Proof*, 55 ARIZ. L. REV. 557, 566–67 (2013) ("[F]requentist probability is of no use. … Courts have no information about the relative frequencies of relevant events.").

[16] *See* A. W. F. EDWARDS, LIKELIHOOD: EXPANDED EDITION 51 (1992) ("[I]n order to apply Bayes' Theorem to hypotheses not generated by a chance set-up, prior probabilities, for which there is no frequency justification, will have to be invented.").

[17] *E.g. Probability*, BLACK'S LAW DICTIONARY (10th ed. 2014) ("Something that is likely … The degree to which something is likely to occur … The quality, state, or condition of being more likely to happen or to have happened than not …"). In fact, the inventor of the concept of likelihood, Ronald A. Fisher, introduced it in an effort to clarify linguistic confusion among statisticians. R. A. Fisher, *On the Mathematical Foundations of Theoretical Statistics*, 222 PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON, SERIES A 309, 309–11 (1922).

Bayesian posterior probability depends critically on prior beliefs, likelihood is a belief-agnostic statement about the consistency of evidence with contrasting hypotheses about the world.

In formal notation, a likelihood function is represented as follows:

$$L(\theta; x) = c \times P(x|\theta)$$

where $c > 0$ can be any arbitrary constant.[18] The argument of interest in the likelihood function, θ, is a variable that takes on all the possible hypotheses of interest. For example, in the cup-choice puzzle, θ has two possible values: (1) "white cup chosen," and (2) "red cup chosen." The conditioning variable in a likelihood function, x, is the evidence that has been observed. In the cup-choice puzzle, this is the white marble drawn at random from the unidentified cup.

Likelihoods are related to probability statements by definition. But likelihoods are not probabilities.[19] At a deep level, probabilities and likelihoods behave differently.

- Probabilities have individual meaning (a ½ probability of an event is an absolute statement of frequency or belief in that proposition); likelihoods have no individual meaning, and can only be interpreted in comparison to other likelihoods.[20]

[18] The point of the constant, $c > 0$, is to act as a reminder that likelihood has no meaning on its own. A likelihood of $L(A) = 8$ says nothing about the probability of A; it only reflects that A is twice as likely as B if $L(B) = 4$, since $L(A)/L(B) = 2$. This ratio is the same for any scaling constant, $c > 0$.

[19] *See, e.g.*, YUDI PAWITAN, IN ALL LIKELIHOOD: STATISTICAL MODELING AND INFERENCE USING LIKELIHOOD 17 (2013) ("The fundamental difference is that *the likelihood does not obey probability laws*. So probability and likelihood are different concepts available to deal with different levels of uncertainty.") (emphasis in original).

[20] *E.g. id.* at 207 ("[A] likelihood ratio compares the relative merits of two hypotheses in light of the data; it does not provide an absolute support for or against a particular hypothesis on its own."); RICHARD ROYALL, STATISTICAL EVIDENCE: A LIKELIHOOD PARADIGM 24 (1997) (making a similar observation); *see also supra* note 18.

- Probabilities add to one over all possible values of the random variable; likelihoods need not add to one (or even a finite number) over all possible values of their variable of interest.[21]
- Probabilities allow simple hypotheses, like $P(A)$, to be compared to composite hypotheses, like $P(B \text{ or } C)$; likelihoods can generally only be used to compare two simple hypotheses at a time.[22]

It may sound like likelihoods lack many of the attractive properties of probabilities—and they do. Compared to probability, likelihood is a weaker (less descriptive) concept of uncertainty.[23] But likelihoods have their own attractive properties, owing to the special features of likelihood ratios. Of particular importance is the *Law of Likelihood*, a variety of related arguments that, for a given probability model, observed evidence, x, favors hypothesis A over hypothesis B *if and only if* the likelihood of A given x exceeds the likelihood of B given x:

$$LR = \frac{L(A;x)}{L(B;x)} > 1$$

The Law further provides that the strength of this evidential support is entirely represented by the magnitude of the likelihood ratio.[24]

In the cup-choice puzzle, for example, drawing a white marble from the unidentified cup is evidence in favor of the hypothesis that I chose the white cup over the alternative that I chose the red cup:

---

[21] *E.g.* IAN HACKING, LOGIC OF STATISTICAL INFERENCE 50 (2016) (noting that likelihoods do not obey the Kolmogoroff axioms); Fisher, *supra* note 17, at 327 (similar).
[22] *E.g.* Royall, *supra* note 20, at 16–20 (explaining and detailing this property).
[23] *See, e.g.*, Pawitan, *supra* note 19, at 15 (noting that Fisher considered likelihood weaker than probability, appropriate to "analyze [and] communicate statistical evidence of types too weak to supply true probability statements") (citing RONALD A. FISHER, STATISTICAL METHODS AND SCIENTIFIC INFERENCE 73–75 (3d ed. 1973)).
[24] This description of the *Law of likelihood* is based on Ian Hacking's formulation. Hacking, *supra* note 21, at 48–66; *see also* Edwards, *supra* note 16, at 28–31 (similar); Royall, *supra* note 20, at 1–3 (similar). For a formal argument for the related likelihood principle, see Allan Birnbaum, *On the Foundations of Statistical Inference* 57 J. OF THE AM. STAT. ASS'N 269 (1962).

$$\frac{L(white\ cup\ chosen;\ white\ marble\ drawn)}{L(red\ cup\ chosen;\ white\ marble\ drawn)} = \frac{c \times 5/6}{c \times 1/6} = 5$$

The likelihood-ratio of 5 means that the evidence is five times more consistent with the white-cup hypothesis than it is with the red-cup hypothesis. This ratio would have the same interpretation in a medical context, a physical science context, or a legal fact-finding context.[25] Likelihood ratios are a unitless measure of the relative consistency of competing hypotheses with observed evidence.

But what a likelihood ratio of 5 emphatically does not say is that the *probability* of A is 5 times the probability of B. These hypotheses may concern historic acts or fixed constants, and one need not know or be able to say *anything* about the prior probabilities of events to use likelihood reasoning. Likelihood analysis of the cup-choice puzzle thus imposed no restriction at all on my choice of cup. In fact, this reasoning works just as well even if *no prior probability distribution c*ould describe that choice.[26] This is the beauty of likelihood reasoning: on the basis of evidence—and evidence alone—likelihood "supplies a natural order of preference among the possibilities under consideration."[27]

The connection to fact-finding is again hard to miss: likelihood provides an independent language for describing fact-finding. If certain evidence is more likely to arise under one set of facts than another, then its observation supports the more likely factual hypothesis, and the magnitude of that likelihood ratio reflects the strength of this support. In fact, the likelihood ratio arguably summarizes *all* of the information contained in the entire body of observed evidence.[28]

---

[25] *See* Royall, *supra* note 20, at 27.
[26] *See* Edwards, *supra* note 16, at 55–65 (discussing the impossibility of representing ignorance by a prior probability distribution); Royall, *supra* note 20, at 173 (same).
[27] RONALD A. FISHER, STATISTICAL METHODS AND SCIENTIFIC INFERENCE 68 (1956) (quoted in Edwards, *supra* note 16, at 27).
[28] *See, e.g.*, Edwards, *supra* note 16, at 30 ("Within the framework of a statistical model, *all* the information which the data provide concerning the relative merits of two hypotheses is contained in the likelihood ratio of those hypotheses on the data.") (emphasis in original).

2.3 *Persuasion (Fact-Finding under Uncertainty)*

Like both probability and likelihood, persuasion is a concept arising from uncertainty. If fact-finders could divine the facts of a case without error or doubt, then persuasion would never enter the picture: the facts would simply be compared to the cause of action to resolve each dispute. The world is, of course, not so perfect. Material facts are often still uncertain at the close of evidence,[29] and so burdens of persuasion are used to categorize evidence that is *good enough* to legally establish yet uncertain facts, claims, and defenses.

So if probability describes belief, and likelihood describes evidential support, which is the measure of persuasion? It would be nice if courts or lawmakers had ever consciously adopted one measure or the other. But they haven't. And common articulations of every burden of persuasion are now a muddled mess of imprecise notions of probability, likelihood, certainty, doubt, and the weight of evidence.

For example, in most civil actions the burden of persuasion is proof *by preponderance of the evidence*. This is often explained as requiring the fact-finder to "be persuaded by the evidence that the claim [or affirmative defense] is more probably true than not true."[30] But it is also explained as requiring "any underlying material fact [to be] more likely [true] than not."[31] As well as requiring that "the scales tip, however slightly, in favor of the party with [the] burden [of persuasion]."[32]

In most criminal actions, due process demands a more stringent standard: proof *beyond a reasonable doubt*.[33] Explanations of this bur-

---

[29] *See* McBaine, *supra* note 1, at 246 ("[C]ertainty as to what has happened cannot be ascertained from the testimony of witnesses or other evidence of acts. The frailty of man is such that certainty in the field of fact finding is impossible.").

[30] Ninth Circuit Manual of Model Jury Instructions: Civil § 1.13 (2007) (bracketed text in original), available at http://www3.ce9.uscourts.gov/jury-instructions/sites/default/files/WPD/Civil_Instructions_2016_6.pdf.

[31] Aguilar v. Atl. Richfield Co., 24 P.3d 493, 507 (2001); *see also* Guglielmino v. McKee Foods Corp., 506 F.3d 696, 698 (9th Cir. 2007) (defining the preponderance standard as that in which a claim is shown "more likely than not").

[32] Ostrowski v. Atlantic Mut. Ins. Cos., 968 F.2d 171, 187 (2d Cir.1992).

[33] *See* In re Winship, 397 U.S. 358, 364 (1970) (holding "that the Due Process Clause protects the accused against conviction except upon proof beyond a reasonable doubt of every fact necessary to constitute the crime with which he is charged.").

den rarely use either probability or likelihood language. Instead, proof beyond a reasonable doubt is described as something like "proof that leaves you firmly convinced [that] the defendant is guilty."[34] Or in the negative, it is said to be lack of a doubt which "would cause a prudent man to hesitate in taking action upon an important matter," as opposed to only "arbitrary" or "speculative" doubt.[35]

In between these extremes, some civil claims and defenses are held to intermediate standards like proof *by clear-and-convincing evidence*. These intermediate burdens are not as standardized as the main two, nor as well understood.[36] And that is failing to meet a low bar, since confusion about even the main two burdens of persuasion has embarrassed the legal system for generations.[37]

Centuries of efforts to clarify the burdens of persuasion with linguistic refinements having failed,[38] the past 50 years have instead seen many courts and legal scholars turn to probability ideas in their efforts to say what these standards mean. Edward Cheng recently summarized the probability thresholds that are now often presented as an explanation of the common burdens of persuasion:

> 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. By comparison, the criminal [reasonable-doubt] standard is akin to a probability greater than 0.9 or 0.95.[39]

---

[34] Ninth Circuit Manual of Model Jury Instructions: Criminal § 3.5 (2010), available at http://www.rid.uscourts.gov/menu/judges/jurycharges/OtherPJI/9th%20Circuit%20Model%20Criminal%20Jury%20Instructions.pdf.

[35] McBaine, *supra* note 1, at 257 (summarizing these and other expressions).

[36] *See* 2 McCormick on Evidence § 340 (Kenneth S. Broun et al. eds., 7th ed. 2013) (citing additional examples such as "clear, convincing and satisfactory," "clear, cogent, and convincing" and "clear, unequivocal, satisfactory and convincing [evidence]" of these standards, and commenting that "[n]o high degree of precision can be attained by these groups of adjectives").

[37] *See supra* note 1 and accompanying text.

[38] McCormick on Evidence, *supra* note 36, at 662–63 (commenting on wasted effort arguing over linguistic metaphysics and "word-magic").

[39] Cheng, *supra* note 3 at 1256.

These thresholds are expressed in terms of the probability of a single proposition. As ratios of probabilities (the probability of guilt versus innocence, for example) the thresholds would be more like 1.0 for preponderance and 9.0–19.0 for reasonable doubt, but the idea is the same.[40] Intermediate standards are less frequently quantified, and presumably fall somewhere between these extremes.

Many scholars harbor serious doubts about the Bayesian approach to fact-finding,[41] but with no alternative theory to take its place, the probability approach has long dominated academic research.[42] Courts are slowly adopting this approach too. While most continue to resist quantifying the reasonable doubt standard,[43] some now explicitly endorse a 50% probability threshold when explaining preponderance of the evidence.[44] And surveys of judges show rough consensus that preponderance of the evidence requires greater than a 50–60% probability of a fact's truth,[45] while reasonable doubt requires something higher, like a 75–90% probability of truth.[46]

---

[40] For example, a 0.9 probability of guilt, implying a $1.0 - 0.9 = 0.1$ probability of innocence, corresponds to a $0.9/0.1 = 9.0$ probability ratio of guilt to innocence.

[41] *See generally*, Allen, *supra* note 5; Allen & Stein, *supra* note 15.

[42] *See, e.g.*, Gary L. Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?* 62 J. PERSONALITY & SOC. PSYCHOL. 739, 739 (1992) (noting that "the probability-threshold model" is "the dominant decision model put forth in the [literature]" and citing decades of research on this model).

[43] *E.g.* Commonwealth v. Sullivan, 20 Mass. App. Ct. 802, 804–05 (Mass. App. Ct. 1985) (finding error in trial court's illustrative quantification of reasonable doubt standard in probability terms in response to jury request for clarification).

[44] *E.g.* Brown v. Bowen, 847 F.2d 342, 345–46 (7th Cir. 1988) ("All burdens of persuasion deal with probabilities. The preponderance standard is a more-likely-than-not rule, under which the trier of fact rules for the plaintiff if it thinks the chance greater than 0.5 that the plaintiff is in the right. The reasonable doubt standard is much higher, perhaps 0.9 or better. The clear-and-convincing standard is somewhere in between.").

[45] *See, e.g.*, U.S. v. Fatico, 458 F.Supp. 388, 409-10 (1978) (citing and surveying evidence from two such surveys).

[46] *See, e.g.*, Posner, *supra* note 14, at 1506 (surveying studies in noting that "Judges, when asked to express proof beyond a reasonable doubt as a probability of guilt, generally pick a number between .75 and .90"). *See generally*, Reid Hastie, *Algebraic Models of Juror Decision Processes*, *in* INSIDE THE JUROR: THE PSYCHOLOGY OF JUROR DECISION MAKING 192 (Reid Hastie ed., 1993) (summarizing many studies eliciting probability thresholds for the preponderance and reasonable doubt standards).

Likelihood reasoning is not entirely missing from the conversation, but it is rarely treated as more than a steppingstone on the path to the posterior probabilities of interest, and never considered an independent concept of uncertainty for the fact-finding process. The closest the literature has come to a likelihood theory of fact-finding is in recent papers by Louis Kaplow and Edward Cheng. In 2012, Kaplow proposed to implement a new system of fact-finding around "evidence thresholds" derived from assumptions about the social objectives of the justice system. Kaplow's new system of fact-finding would use a form of likelihood reasoning, but his reading of current fact-finding and burdens of persuasion is consistent with the usual Bayesian probability approach.[47] In 2013, Cheng proposed to model fact-finding in terms of posterior probability ratios, but under *ad hoc* assumptions that happened to make these Bayesian probabilities behave like likelihoods in many respects.[48] And in 2014, Kaplow again suggested that likelihood ratios could be used to model various legal decision-making problems, but again interpreted current burdens of persuasion in terms equivalent to the traditional Bayesian approach.[49] Suffice it to say that while recent work is skirting the edges of a likelihood theory of fact-finding, the critical step of actually deriving a pure theory of fact-finding from likelihood reasoning alone has yet to be undertaken. The next section does so.

---

[47] Kaplow, *supra* note 9, at 748 (distinguishing current law, which "takes behavior as given … and asks, in light of that behavior, what is the likelihood of [harmful or benign] acts" from the proposed "welfare-based, optimal threshold" approach in which the central question is "how behavior … will change as a function of a change in the evidence threshold."). Care is needed in interpretation, however, as Kaplow does not always distinguish probability from likelihood in his terminology. *See, e.g.*, *id.* at 758-59; *id.* at 748 n.19.

[48] Cheng, *supra* note 3, at 1263–65 (assuming probabilities of alternative fact-combinations cannot be aggregated in fact-finding); *id.* at 1267–68 and nn.24–25 (assuming the ratio of prior probabilities must always equal 1 in a legal setting).

[49] Louis Kaplow, *Likelihood Ratio Tests and Legal Decision Rules*, 16 AM. L. ECON. REV. 1, 35 (2014) ("[Preponderance of the evidence] can be equivalently stated in terms of the posterior probabilities … or in terms of the likelihood ratio…").

3 A LIKELIHOOD THEORY OF FACT-FINDING

One cannot think clearly about fact-finding without the foundation of a framework for how facts, evidence, and causes of action interact in the justice system. Yet research on fact-finding often does not stop to construct such a framework,[50] which may itself explain why likelihood reasoning is not more prevalent today. As the following shows, even a cursory effort to build the underlying framework leads immediately to a likelihood theory of fact-finding.

3.1 *The General Fact-Finding Framework*

Distilled down to core concepts, legal fact-finders are tasked with comparing uncertain facts to the elements of a cause of action. They do so by reference to evidence put forth by litigants, and subject to whatever burden of persuasion represents sufficient evidence to prove the facts in a given context. None of this is new ground, but it still helps to take the uncommon step of considering the properties of each component in this framework.

Facts are the actions, omissions, intents, and beliefs of the parties that are material to a claim for legal relief. These facts could include some random elements (like accidental injury resulting from the defendant's negligence) but will more often consist of historic acts and choices (like the defendant's conscious election not to undertake certain safety measures). Mixed questions of fact and law (whether failure to undertake these safety measures breached a duty of care) are themselves simply functions of more basic facts (what measures could have been taken; what would their efficacy have been), and are no different from standard facts at a theoretical level.

A *cause of action* is a set of facts sufficient to warrant legal relief. If F denotes the universe of all combinations of material facts that could plausibly be true, then let $C \subset F$ denote the subset of that universe in

[50] *But cf.* David H. Kaye, *Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt?*, 66 B.U. L. REV. 657, 659–61 (1986) (distinguishing between facts, evidence, and narrative stories in a manner similar to the following).

which the combination of facts makes out a cause of action.[51] In a negligence action, for example, C would be all combinations of facts that satisfy the elements of duty, breach of duty, causation, and damages; in a criminal action, it would be all combinations of facts that fit the elements of the crime charged. If a given combination of facts falls within the cause-of-action set, $f \in C$, then it justifies legal relief. The complement of the cause-of-action set, C^c, is the set of all possible combinations of facts in F that are not in C. Any combination of facts falling in this no-remedy set, $f \in C^c$, fails to justify relief.

Evidence is what the fact-finder sees and hears at trial. It is sometimes forgotten that the evidence is not the facts. It is not usually even direct proof of the facts. At trial, the fact-finder hears testimony and sees documentary evidence that bears—in totality—on what the facts might be. Some of this evidence is direct proof of the facts. Some of it is evidence explaining why other evidence has or hasn't been shown: impeachment of a witness by bias or prior inconsistent statement,[52] presence or absence of records of a regularly conducted activity,[53] etc. Some of it is evidence presented to build a broader narrative or story.[54] From this body of disparate and usually conflicting evidence, the fact-finder is asked to draw inferences about the facts of the case.

Unlike the facts, it will generally make sense to think of the body of evidence as generated by a random process. To see why, consider even a simple negligence action in which the motivating injury was seen by many bystanders. As a practical matter, neither party can force a given

---

[51] The set notation in this paragraph is basic, and describes relationships succinctly. The notation $x \in A$ means "x is an element of the set A." For example, if A is a set with two elements, $A = \{1, 4\}$, then $4 \in A$ but $3 \notin A$. The notation $A \subset B$ means "A is a subset of B." For example, if $A = \{1, 4\}$ and $B = \{1, 4, 5\}$ then $A \subset B$ because B contains both 1 and 4, but $B \not\subset A$ because A is missing 5.

[52] *E.g.* United States v. Abel, 469 U.S. 45, 52 (1984) (explaining bias); FED. R. EVID. 607 (attacking credibility); FED. R. EVID. 613 (prior inconsistent statement).

[53] *E.g.* FED. R. EVID. 803(6)–(7) (presence and absence of certain records).

[54] *E.g.* State v. Villavicencio, 95 Ariz. 199, 201 (Ariz. 1964) ("[The] principle that the complete story of the crime may be shown even though it reveals other crimes has often been termed 'res gestae'… we choose to refer to this as the 'complete story' principle."); *see also* Kaye, *supra* note 50, at 662–65 (discussing the role of stories and "gaps" in evidence in a probability theory of fact-finding).

bystander to take the stand as a friendly witness: cooperation is largely luck-of-the-draw. And even if a witness is willing to cooperate, her testimony will only be as good as her memory and communication skills permit. Finally, her potential testimony will only reach the fact-finder if the lawyers remember to introduce it, and the rules of evidence allow.[55] The same goes for documentary evidence: documents may not be retained long enough to be discovered, may be retained but still not discovered, may be discovered but not reviewed in a large production, may be reviewed but not properly interpreted, may be properly interpreted but impossible to authenticate,[56] may be possible to authenticate but difficult to read or present, and so on. In sum, the particular body of evidence that ultimately gets on the record is the result of a long and complicated process over which neither party has great control—a process with a large random component.

But this randomness does not mean that the evidence is unrelated to the facts. To the contrary, if the plaintiff really did suffer an injury as the result of the defendant's negligence, it seems probable that at least one of the bystanders would be willing to take the stand to testify to that effect. And, similarly, the more negligent the defendant's acts, the less probable it is that she would be able to find a bystander whose imperfect recollection painted her as prudent. Put formally, the probability distribution generating the evidence depends on the underlying facts.[57] So if E is the universe of all evidence that could possibly make the record, then the probability of observing a given combination of evidence, $e \in E$, will usually vary with the hypothesized facts: $P(e|f)$.

For purposes of laying out the likelihood theory itself, the existence of these conditional probabilities, $P(e|f)$, can be taken as primitive:

[55] *E.g.* Federal Rule of Evidence 801–807 (hearsay); 601–602, 701 (competence); 403 (undue prejudice etc).

[56] *E.g.* FED. R. EVID. 901.

[57] *See, e.g.*, Allen & Stein, *supra* note 15, at 577 ("Virtually always, therefore, this evidence will have some causal connection to the story's truth. To put it differently, this evidence would not have come into existence the way it did had the story been false rather than true."); Lempert, *supra* note 13, at 1052 ("Upon hearing testimony, jurors must compare the probability that the testimony *would be given* if the defendant were guilty with the probability that the testimony *would be given if* the defendant were innocent.").

something the fact-finder knows or can assess. This assumption is basic to a Bayesian probability approach to fact-finding as well,[58] and is actually more restrictive in that context than it is here. Likelihood reasoning does not depend on the absolute probability of the evidence or anything else; simple relative probabilities are entirely sufficient.

In trial fact-finding, the practical analogy to this type of likelihood reasoning is described by Ron Allen and many others in their work on the cognitive process of fact-finders.[59] To oversimplify that work, fact-finders assess the merits of competing factual narratives by comparing their epistemic credentials on the observed evidence. Factors considered in this process include the coherence, consilience, causal specificity of stories about the facts, and other assessments of the comparative probability of observing the evidence under different hypotheses.[60] Put bluntly, this empirical research on fact-finder cognition suggests that fact-finders interpret evidence in relation to factual narratives in a way that is hard to distinguish from likelihood reasoning.

3.2 *The Likelihood Approach to Fact-Finding*

An elegant theory of legal fact-finding, based on likelihood reasoning alone, can be used to interpret and explain every burden of persuasion in use today. All this theory requires is the grounding of the foregoing fact-finding framework, and the special properties of likelihood ratios. The theory closely fits much of the language, practice, and intuition of

---

[58] This assumption has undergirded the Bayesian analysis of fact-finding since the earliest works on this topic. *E.g.* Kaplan, supra note 13, at 1085 ("[T]he [fact-finder] will … determine as best he can … the probability that the piece of evidence would have occurred under the hypothesis of guilt to the probability that it would have occurred under the hypothesis of innocence.").

[59] *See, e.g.*, *supra* notes 5–7; *see also* Allen & Stein, *supra* note 15, at 567–71 (discussing and citing literature related to the relative plausibility model).

[60] *See* Allen & Stein, *supra* note 15, at 577 ("[E]vidence that allows the winning story to win … does not come into existence by accident. This evidence must satisfy a demanding set of epistemic criteria [such as coherence, causal specification, evidential support, and other criteria associated with natural reasoning]. Virtually always, therefore, this evidence will have some causal connection to the story's truth. To put it differently, this evidence would not have come into existence the way it did had the story been false rather than true.").

current legal fact-finding.

The easiest way to explain this theory is to start with how it defines the ultimate fact-finding inquiry. Every burden of persuasion in use today can be reduced to the same rule of likelihood reasoning: find for the plaintiff *if and only if*

$$LR = \frac{\sup_{f \in C} L(f; e)}{\sup_{f \in C^c} L(f; e)} > k$$

where $k \geq 1$ is a threshold value determined by the applicable burden of persuasion. Let me explain.

Broken down, this test has two parts: to the left of the inequality is a likelihood ratio; to the right, a threshold value. The reliance on likelihood ratios shouldn't be surprising. By the Law of Likelihood, the observed value of a random variable (evidence) is evidential support for one hypothesis (factual theory) over another (opposing factual theory) *if and only if* the respective likelihood ratio is greater than one. The point of the threshold value is also explained by the Law of Likelihood. Since the magnitude of a likelihood ratio conveys how strongly the evidence favors the top hypothesis over the bottom, larger values of k correspond to increasingly demanding evidentiary requirements for the top hypothesis to be accepted—e.g., for the plaintiff to prevail.

The *sup* (supremum) terms in the likelihood ratio instruct the fact-finder to select the *most* likely combinations of facts in two mutually exclusive and exhaustive subsets of the universe of possible facts. The top term is the likelihood of the most plausible combination of facts in the cause-of-action set: the likelihood of the most likely pro-plaintiff factual theory. The bottom term is the same, but for the no-remedy set: the likelihood of the most likely pro-defendant factual theory. Put simply, this ratio compares the likelihoods of the single *most* plausible factual theory favoring each side of the case.

This ultimate comparison of only the most likely factual hypothesis favoring each side of the case is intuitive, and aligns with the typical use of likelihood ratios in statistics.[61] The comparison makes particular sense in the trial context, where for both sides advocate the relative plausibility of their respective factual theories. We might suppose that fact-finders will often end up simply comparing the parties' own factual theories, but nothing prevents some other theory from appearing more plausible during the course of trial.[62]

Some may find it surprising that the final analysis is limited to only the two most likely factual theories, but on both technical and intuitive grounds it actually makes a great deal of sense. First, to be clear, this framework does *not* prevent either party from arguing in the alternative; it simply means that alternative arguments will be considered individually, not in aggregate. Second, the non-aggregation of alternative hypotheses is not artificial or imposed on the framework, but a substantive implication of the non-existence (or disregard) of prior probabilities in likelihood analysis. Likelihood is a more credible basis for inference than Bayesian analysis when prior probabilities are artificial and subjective; but this same avoidance of prior probabilities renders likelihood a weaker form of reasoning. In particular, aggregation of confidence across several alternative hypotheses is generally not possible when reasoning in likelihood terms.[63] Third, this ultimate choice between two fully specified factual theories is an intuitively desirable property for *any* theory of legal fact-finding. It means that fact-finding always results in a single, specific finding of fact—something that an approach based on the aggregation of alternative factual hypotheses could not promise.

Finally, the likelihood ratio comparing the plausibility of these two factual theories is itself compared to a threshold value derived from the burden of persuasion. As explained in the following subsections,

[61] The comparison of suprema of likelihoods within different categories of hypotheses is how likelihood ratio tests are typically conducted in statistics. *See generally* GEORGE CASELLA & ROGER L. BERGER, STATISTICAL INFERENCE 373–79, 385–91 (2d ed. 2002) (discussing the construction of statistical likelihood ratio tests).
[62] Allen & Stein, *supra* note 15, at 568 ("[Fact-finders] consider the parties' competing stories and decide which is superior; in some cases, they construct their own account of the events in light of the parties' evidence and arguments.").
[63] *See supra* notes 19–22 and accompanying text.

that threshold values is $k = 1$ for preponderance of the evidence, $1 < k < 10$ for clear-and-convincing evidence, and $k \geq 10$ for the reasonable-doubt standard. The preponderance threshold is clear, but the two heightened-burden thresholds are theoretically ambiguous, and would require future empirical refinement to operationalize.

3.2.1 THE PREPONDERANCE-OF-THE-EVIDENCE THRESHOLD (K=1)

Deriving the appropriate threshold value for the preponderance standard—and every other standard, for that matter—requires inquiry into the underlying social and legal objectives behind the burden. Much commentary on the preponderance standard is an unhelpful mess of verbal gymnastics[64] and imprecise notions of evidentiary weight.[65] But from this general confusion, two clear principles do emerge.

First, the preponderance standard places no special weight on the direction of any mistakes the fact-finder might make. There is near universal agreement that, at least in the usual civil suit, the risk of an erroneous factual finding falls no heavier on the plaintiff than it does on the defendant. Justice Harlan provides a typical statement of this normative view in a well-known concurrence:

> In a civil suit between two private parties … we view it as no more serious 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.[66]

Similar assertions are often made by scholars commenting on the basis for the preponderance standard.[67]

---

[64] *See supra* note 38.

[65] *See, e.g.*, MCCORMICK ON EVIDENCE, *supra* note 36, at 660 n.7 (noting common judicial instruction against simply adding up the number of witnesses).

[66] In re Winship, 397 U.S. 358, 371–72 (1970) (Harlan, J., concurring).

[67] *E.g.* Mike Redmayne, *Standards of Proof in Civil Litigation*, 62 MODERN L. REV. 167, 171 (1999) ("[In a 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."); MCCORMICK ON EVIDENCE, *supra* note 36, at 669 (making a similar assertion).

Second, no particular strength of evidence is needed to meet this burden. Sometimes described as the "greater weight of the evidence,"[68] sometimes as facts just "more likely than not" in the plaintiff's favor,[69] the inquiry is consistently into the *direction* that the evidence points, and not into how strongly it points that direction.[70]

These two principles translate to likelihood reasoning as a threshold value of $k = 1$. Demanding that the likelihood ratio exceed $k = 1$ for the plaintiff to prevail is a literal translation of the requirement that the scales of evidence tip, however slightly, in the plaintiff's favor:[71] any value of the likelihood ratio greater than one means that the weight of the evidence favors the plaintiff. The $k = 1$ threshold value also respects the premise of Justice Harlan, and others, that the preponderance standard should treat the plaintiff and defendant symmetrically. Any other choice of threshold would have the effect of according extra weight to the candidate factual theory of one party or the other; only $k = 1$ affords both sides equal weight in the final inquiry.

An interesting corollary of this threshold value is that it reduces the fact-finding inquiry to a simple search for the most likely factual story on the evidence. Since the most likely factual story is necessarily the candidate theory favoring one of the two parties,[72] the winning story under the preponderance standard is, by definition, the *most* likely factual story on the evidence.[73] Put another way, this likelihood theory reduces the preponderance-of-the-evidence standard to a simple question: what is the *most* plausible combination of facts on the observed evidence? The party it favors wins.

[68] *E.g.* United States v. Matlock, 415 US 164, 178 n.14 (1974).
[69] *E.g.* St. Mary's Honor Center v. Hicks, 509 US 502, 527–28 (1993).
[70] *But cf.* McBaine, *supra* note 1, at 238–39 (noting that some courts that have tried to distinguish relative weight-of-evidence from absolute persuasion in this context).
[71] *Supra* note 32.
[72] Since C and C^c exhaust all possible combinations of facts in the universe, the most likely combination of facts on the evidence must be either the most likely combination of facts in C or the most likely combination of facts in C^c.
[73] This assumes unique suprema. Otherwise multiple factual stories might be equally most likely, which complicates discussion of the likelihood theory, but little else.

3.2.2 THE CLEAR-AND-CONVINCING EVIDENCE THRESHOLD ($1 < k < 10$)

Intermediate burdens like clear-and-convincing evidence are more a class of standards than a single entity. For different reasons including social policy, judicial confidence, and special stakes, certain claims and defenses require more than a mere preponderance of the evidence to prevail.[74] While asymmetric error tolerance is implicit in these contexts,[75] it is less emphasized than the heightened evidentiary requirement placed on the party bearing the burden.[76]

Building a strength-of-evidence requirement into the likelihood approach is trivial. Any threshold value $k > 1$ requires the evidence to be at least *k times more likely* under the candidate pro-plaintiff factual theory for the plaintiff to prevail. Intuitively, this splits the strength of evidence into three categories: evidence that does not favor the plaintiff ($LR \leq 1$), evidence that weakly favors the plaintiff ($1 < LR \leq k$), and evidence that strongly favors the plaintiff ($LR > k$). The first two categories both require finding for the defendant; only *strong evidence* suffices to justify relief under these intermediate burdens.

None of this says what the numerical value of k is—and it may well differ by type of case and jurisdiction.[77] The proper threshold value is ultimately an empirical question, which I cannot answer from theory alone. But in seeking an empirical estimate for this value, numerical analogies are available to assist. In the cup-choice puzzle, for example, the evidence of a white marble drawn at random from the unidentified

---

[74] *See, e.g.*, MCCORMICK ON EVIDENCE, *supra* note 36, at 665 (noting intermediate standards in a "variety of cases involving deprivations of individual rights not rising to level of criminal prosecution"); *id.* at 665–66 (noting intermediate standards that reflect inherited rules of fact-finding from courts of equity); *id.* at 668 (noting intermediate standards for claims "disfavored on policy grounds").

[75] *See, e.g.*, Colorado v. New Mexico, 467 US 310 (1984) (adopting a clear-and-convincing-evidence requirement to dictate who "should bear most, though not all, of the risks of erroneous decision" but also to "accommodates society's competing interests in increasing the stability of property rights and in putting resources to their most efficient uses").

[76] *E.g.* MCCORMICK ON EVIDENCE, *supra* note 36, at 659–60 (commenting that unlike reasonable doubt, the preponderance and clear-and-convincing formulations direct attention to the evidence); McBaine, *supra* note 1, at 253 (commenting similarly).

[77] *See supra* note 36.

cup would be strong enough evidence to satisfy any threshold value, $k < 5$. If drawing a white marble in this context feels like clear-and-convincing evidence that I chose the white cup, then this is one data point in favor of a threshold value of at most 5. If this evidence does not feel clear and convincing, then the exercise may be modified to increase the strength of evidence until it does. There are many exercises to try in this type of calibration process,[78] and in future research rough consensus might be hammered out on an appropriate threshold value. For now, it can only be given as a range.

3.2.3 THE REASONABLE-DOUBT THRESHOLD ($K \geq 10$)

It has often been said that the reasonable-doubt standard differs from the preponderance and clear-and-convincing evidence standards: where the latter two focus on the weight of the evidence, the former seems to look to the mental state of the fact-finder.[79] This seems more a point of form than substance. The only clear social and legal distinction between reasonable doubt and these other standards is the special concern about false convictions in the criminal context. Justice Harlan makes this point explicit in the same concurrence as before:

> I view the requirement of proof beyond a reasonable doubt in a criminal case as bottomed on a fundamental value determination of our society that it is far worse to convict an innocent man than to let a guilty man go free.[80]

Similar statements have motivated the Court's decisions in other cases involving this standard,[81] and the overarching false-conviction concern

---

[78] *See, e.g.*, Royall, *supra* note 20, at ch. 1 (providing several examples).
[79] *See, e.g.*, McBaine, *supra* note 1, at 255 ("In criminal cases the extent or degree of belief of the triers of the fact is stressed, not the amount or quality of evidence."); MCCORMICK ON EVIDENCE, *supra* note 36, at 659–60 (making a similar observation).
[80] In re Winship, 397 U.S. 358, 372 (1970) (Harlan, J., concurring).
[81] *E.g.* Addington v. Texas, 441 US 418, 423–24 (1979) ("In a criminal case … the interests of the defendant are of such magnitude that … they have been protected by standards of proof designed to exclude as nearly as possible the likelihood of an er-

is widely endorsed by the broader legal community as well.[82]

At first blush, a focus on false convictions would seem problematic for a likelihood theory of fact-finding. After all, likelihood measures strength-of-evidence, not probability-of-mistake. But there is also an intuitive connection between these concepts. We have a general sense that we are more apt to err when acting on weak evidence than when acting on strong evidence—that as evidence becomes overwhelmingly one-sided, the possibility that we are seeing such strong evidence by chance alone becomes less and less plausible. This intuition is born out in likelihood reasoning, particularly in a general bound that can be placed on the probability of observing strong *and* misleading evidence in a given comparison. When contrasting any two factual hypotheses, f' and f'',[83] where f'' is actually true, the probability of seeing strong evidence that misleadingly favors f' over f'' is bounded from above by the inverse of the strength of the evidence:[84]

$$P\left(\frac{L(f';e)}{L(f'';e)} \geq k \;\middle|\; f''\ true\right) \leq \frac{1}{k}$$

Put another way, the long-run probability that this likelihood ratio would spuriously favor f' over f'' cannot exceed $1/k$.

This is not exactly a statement about controlling the global probability of false conviction (because it cannot always be assumed that either f' or f'' is true in a given context),[85] but it does formalize the

roneous [conviction]. In the administration of criminal justice, our society imposes almost the entire risk of error upon itself").

[82] *E.g.* MCCORMICK ON EVIDENCE, *supra* note 36, at 669 ("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 … [for the] worthy goal of decreasing the number of one kind of mistake—conviction of the innocent.").

[83] The notation f' and f'' is read "f prime" and "f double-prime." This is simply a shorthand way of denoting two different factual hypotheses.

[84] Royall, *supra* note 20, at 7.

[85] *See* Mark L. Taper and Subhash R. Lele, *Evidence, Evidence Functions, and Error Probabilities*, *in* 7 HANDBOOK OF THE PHILOSOPHY OF SCIENCE 522 *et seq.* (Malcolm R. Forster & Prasanta S. Bandyopadhyay eds., 2011) (discussing the reliability of likelihood ratio inferences when the true parameter may lie outside the contrast).

intuition that the chances of spurious evidence favoring conviction fall away as increasingly rigorous evidence is required to convict.

Defining *error* in the sense of spuriously observing strong evidence for the plaintiff's candidate factual theory when the defendant's theory is really true, the above bound provides a recipe for translating error tolerance into a threshold in the likelihood ratio test. If as a society we want this long-run rate of error to be no greater than 10%, a threshold value of $k = 10$ is sufficient. If we want the probability of error to be lower yet, perhaps 5%, then a value of $k = 20$ is adequate. The long-run rate of error can be made arbitrarily small by demanding an increasingly strong evidentiary showing to convict.

But three caveats are in order. First, these probability bounds are not necessarily tight. So while a threshold value of $k = 10$ guarantees no more than a 10% error rate, the actual frequency of strong and misleading evidence could be substantially lower.[86] Second, this is a very narrow definition of wrongful conviction, in which errors are based on the assumption that either the plaintiff's or defendant's candidate theory is true. That may well be a reasonable approach, given that we cannot know the actual facts by definition, but it makes the interpretation a little different than a true global bound on the rate of false conviction, and the two concepts should not be confused. Third, care is needed in interpreting the error rate. Evidence of strength $k = 10$ does *not* mean that there is at most a 10% probability that the evidence is misleading in a given comparison. Rather, the *process* of using this likelihood ratio test and threshold value would, in the long run, yield no more than 10% strong and misleading comparisons.

3.3 *The Intuition behind the Likelihood Approach*

This likelihood theory of fact-finding reduces every burden of persua-

[86] Tighter bounds (for which lower threshold values could guarantee the same rate of error) may be derived in specific circumstances where more is known about the relevant probability distributions. *See, e.g.*, Royall, *supra* note 20, at 90–94. In a different but related context, Cheng notes that computing the exact probability of false conviction requires very specific information about the probability distributions in question. Cheng, *supra* note 3, at 1277–78.

sion to the same simple rule of likelihood reasoning, and that rule can be further reduced to a four-step algorithm for deciding any case.

- First, locate the most likely set of facts in which the plaintiff makes out a cause of action on the evidence.
- Second, locate the most likely set of facts in which the plaintiff fails to make out a cause of action on the evidence.
- Third, compare the likelihood of these two factual hypotheses.
- Fourth, compare this likelihood ratio to the appropriate threshold value: the plaintiff wins if and only if the ratio exceeds this value.

This theory of the fact-finding process differs fundamentally from the conventional Bayesian probability account of fact-finding. Unlike probability, likelihood is an intrinsically relative concept of evidence.[87] And in a likelihood approach to fact-finding, the fact-finder is never required to form absolute propositional beliefs about any combinations of facts in isolation. Rather, every step in the process involves the simple comparison of alternatives. This explanation of the fact-finding process mirrors arguments in the literature that fact-finding in adversarial litigation involves comparative reasoning—not the formation of absolute beliefs.[88] Intuitively, this likelihood theory formalizes much of the recent cognitive work on fact-finding, which describes fact-finders

---

[87] *See, e.g.*, Edwards, *supra* note 16, at 28 (noting that probability must be used where an absolute degree of belief in a proposition is needed, but that where relative degree of belief is enough, likelihoods are a sufficient measurement); Royall, *supra* note 20, at 8 ("The [law of likelihood] represents a concept of evidence that is essentially relative, one that does not apply to a single hypothesis, taken alone.").

[88] *E.g.* Ronald J. Allen, *A Reconceptualization of Civil Trials*, 66 B.U. L. Rev. 401, 425–28 (1986) (proposing to approach trials as a comparative analysis of two competing accounts); Ronald J. Allen, *The Nature of Juridical Proof*, 13 CARDOZO L. REV. 373, 422 (1991) ("There may be cases where cardinal reasoning works, but the typical case calls for ordinal reasoning."); Cheng, *supra* note 3, at 1259 ("Because the adversarial structure of legal trials promotes jury comparisons of the parties' claims, preponderance is not an absolute probability … [it] is better characterized as a probability ratio, in which the probability of the plaintiffs story of the case is compared with the defendant's story of the case.").

as engaged in the iterative comparison of alternative stories in an effort to find the *relatively* most plausible story on the evidence.[89]

In fact, the proposed likelihood theory can be seen as formalizating much of the relative plausibility, narrative coherency, and story-based models of fact-finding in the literature.[90] This likelihood theory also adds something to these cognitive models. It explains how they scale to heightened burdens of persuasion by demonstrating the existence of a context-independent strength-of-evidence concept. And it helps to justify a supposed limitation of these models: their oft criticized need for evaluation of a holistic factual story, instead of fact-by-fact comparison against the burden of persuasion. Like these cognitive models, the likelihood theory of fact-finding generally demands evaluation of holistic factual stories. The reason is instructive: since the likelihood of any single item of evidence could depend on the entire set of hypothesized facts, there is, in general, no way to evaluate individual facts in isolation under likelihood reasoning.

A final corollary of this reliance on purely comparative reasoning is that the fact-finder must always consider a no-remedy factual theory. This does not limit the defendant's right to hold the plaintiff to her proof; nor does it mean that the defendant must usually put forth a specific theory to prevail.[91] But even if the defendant does not suggest a specific no-remedy theory, the fact-finder must still consider one, by independent inference if nothing else. This may seem surprising at first, but upon reflection it is squarely consistent with the expectation

---

[89] Daniel Shaviro, *Statistical-Probability Evidence and the Appearance of Justice*, 103 HARV. L. REV. 530, 532 (1989) ("[C]ourts should hold in favor of whichever party appears more likely to be correct."); Allen, *supra* note 5, at 609 ("In civil cases, fact finders are to accept the more plausible of the stories advanced by the parties, and in criminal cases they are to accept the state's case only if no plausible story consistent with innocence has been advanced."); Posner, *supra* note 14, at 1513 ("[T]he realistic benchmark for the plaintiff's case is not the null hypothesis but the defendant's case."); Cheng, *supra* note 3.

[90] *See, e.g.*, *supra* notes 5–7.

[91] *Cf.* Cheng, *supra* note 3, at 1262 ("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'.").

that fact-finding should result in a finding of fact. The final step in this likelihood theory is a choice between two fully specified factual theories, and to decide for either party is to find the factual theory favoring that party to be *legally true* under the applicable burden of persuasion.

4 COMPARISON TO PROBABILITY THEORIES

Despite its dominance in academic and judicial thinking, decades of research on the probability approach to fact-finding has yet to produce a satisfying theory,[92] and growing frustration surrounds the paradoxes and impossibilities that this approach entails.[93] Frustration owes to the mistaken belief that only probability can supply a formal theory of fact-finding. But as the previous section shows, the likelihood theory of fact-finding is simple, intuitive, and a reasonable approximation to empirical accounts of fact-finder reasoning in a litigation context. And as this section shows, the likelihood theory also escapes the problems and paradoxes of probability theories of fact-finding.

4.1 *Total Probability Problems*

Even specifying the basic mechanics of a Bayesian probability theory of fact-finding turns out to be surprisingly difficult. The problem is that probability describes the absolute uncertainty in a system, which means that the probabilities of all possible alternatives sum to one. To illustrate the headaches this can causes for a theory of fact-finding, consider even the simple preponderance-of-the-evidence standard.

The traditional probability approach to the preponderance standard requires that the plaintiff prove "the existence of [any material] fact [to be] more probable than its nonexistence."[94] Put another way,

---

[92] *Compare* Allen, *supra* note 88, at 402–03 (noting problems with probability theories of fact-finding in the 1980s) *with* Allen & Stein, *supra* note 15, at 560–65 (noting the same problems and paradoxes in modern probability theories of fact-finding).
[93] *Cf.* Allen & Stein, *supra* note 15, at 560 ("Application of mathematical probability in the courts of law engenders paradoxes and anomalies that are not easy to avoid or explain away. Relative plausibility, on the other hand, faces no such predicaments.").
[94] Flemming James, *Burdens of Proof*, 47 VA. L. REV. 51, 54 (1961); *see also* McBaine, *supra* note 1, at 260–61 (suggesting a similar rule).

the plaintiff must prove that the probability of a set of facts in the cause-of-action set is greater than 50% or—equivalently—greater than the probability of its negation.[95] If $f' \in C$ is a combination of facts supporting relief, then this traditional probability model would have the fact-finder find for the plaintiff if

$$PR = \frac{P(f' \mid e)}{P(not\ f' \mid e)} = \frac{P(f' \mid e)}{1 - P(f' \mid e)} > 1$$

That is, the idealized fact-finder would use Bayes' Theorem to calculate the posterior probability of f' on the evidence, would divide this by one minus that probability (which equates to the probability of all other possible combination of facts) and would compare this ratio to 1 to see if the plaintiff wins.

But this comparison leads to an immediate absurdity. Suppose a car crash occurs on a 25 mph road. The plaintiff's theory, f', includes a claim that the defendant was driving 60 mph at the time of the crash. The negation of this claim is every other speed the defendant could have been driving: this includes still dangerous speeds (59.5 mph), as well as even more dangerous speeds (70 mph). The test outlined above would count the probabilities of both of these strongly pro-plaintiff alternative facts *against* the plaintiff's right to recover!

In fairness, this is a bit of a straw man: no advocate of probability reasoning would endorse the literal application of the traditional test. Instead, they might argue that the plaintiff can allege a *set* of alternative facts to support recovery. For example, perhaps f' includes the composite fact that the defendant was driving "over the speed limit." But this simply shifts the problem to a different place. Now the possibility that the defendant was not speeding, but instead drunk, counts

[95] Ronald Allen, *On the Significance of Batting Averages and Strikeout Totals: A Clarification of the "Naked Statistical Evidence" Debate, the Meaning of "Evidence," and the Requirement of Proof Beyond a Reasonable Doubt*, 65 TUL. L. REV. 1093, 1093 (1991) ("The conventional conception of civil trials involves comparing the probability of a plaintiff's case to its negation."); Cheng, *supra* note 3, at 1254 n.10 ("[C]onventional legal treatments focus on the likelihood ratio between the plaintiffs story being true and the plaintiffs story being false (as opposed to the defendant's story being true).").

against the right to recover. As would the possibility that the plaintiff was asleep-at-the wheel, or distracted by the radio.

The only way that aggregation of alternative theories fully escapes these bizarre results is if the fact-finder considers *every* combination of facts in the *entire* cause-of-action set in assessing the plaintiff's right to recover. This aggregate probability test would then have the fact-finder find for the plaintiff if[96]

$$PR = \frac{\sum_{f \in C} P(f \mid e)}{1 - \sum_{f \in C} P(f \mid e)} > 1$$

That's asking the fact-finder to do a lot: (1) identify every combination of facts in the entire cause-of-action set, (2) compute the probability of each of these combinations of facts using Bayes' Theorem, and (3) add up these possibly infinitely many probabilities to arrive at a single aggregate figure for comparison to the burden of persuasion.

As the basis for a theory of fact-finding, this aggregated probability model has some unattractive properties. First, the task it contemplates seems out of place in a litigation context. When—if ever—would trial map out the absolute probability of every possible combination of facts in entire cause-of-action set?[97] Second, the model pushes the concept of fact-finding past its breaking point. Suppose each of six alternative factual theories has a 10% probability of being true. Together the aggregate probability of these theories is enough to satisfy the preponderance standard. But then what specific combination of facts has the fact-finder actually *found* in this situation?

Some scholars have abandoned all aggregation strategies in favor of probability ratios for individual pairs of factual theories.[98] Cheng provides a recent example: though continuing to interpret fact-finding in terms of Bayesian posterior probabilities, he imposes the *ad hoc* as-

---

[96] The notation $\sum_{a \in A} x(a)$ is a common shorthand signaling the addition of some function, here $x(a)$, over all arguments in a certain set, here $a \in A$.

[97] *Cf.* James, *supra* note 94, at 52 ("[Our adversarial system] frees the judge and jury of responsibility for investigating and presenting facts and arguments, placing that responsibility entirely upon the respective parties …").

[98] Allen, *supra* note 88, at 425–28 (providing what appears to be the first formal suggestion of this approach).

sumption that the parties cannot make arguments in the alternative, so that only individual combinations of facts are comparable in the probability ratio.[99] If the plaintiff and defendant advance factual theories $f' \in C$ and $f'' \in C^c$, respectively, Cheng's proposed probability model would instruct the fact-finder to find for the plaintiff if

$$PR = \frac{P(f' \mid e)}{P(f'' \mid e)} > 1$$

That is, the fact-finder would compare posterior probabilities for *only* these two individual factual hypotheses in determining the winner.

But *ad hoc* modification of the axiomatic properties of probabilities is an odd way to proceed in theory based on probabilities,[100] and here it creates nearly the opposite paradox to the previous models. Suppose the critical fact in a negligence action is whether the defendant ran a red light. The parties argue that the light was red and green, respectively. If the fact-finder decides that the probability the light was red is $P(R|e) = 0.4$ and the probability the light was green is $P(G|e) = 0.3$, then on Cheng's theory the plaintiff wins and the defendant is liable. But since probabilities always sum to one, these findings necessarily imply that the probability of a yellow light is $P(Y|e) = 0.3$. So Cheng's theory would assign liability under circumstances that actually imply the light was most probably *not* red at all: $P(G \text{ } or \text{ } Y|e) = 0.6$.

All of these problems arise from a common source: reliance on absolute belief in a setting where only relative plausibility makes sense. Specifically, all of the problems derive from the rule that probabilities of all possible events must sum to one. Likelihood is an intrinsically relative concept, and does not share this total probability property. Thus, evidence that increases the likelihood of one factual hypothesis may well increase the likelihood of alternative hypotheses as well. And

---

[99] Cheng, *supra* note 3, at 1262 ("The defendant, particularly in a civil case, may not simply be a contrarian.... The defendant may offer multiple possible alternatives, but each of these alternatives will be judged separately, not simultaneously.").
[100] *Cf.* Allen & Stein, *supra* note 15, at 596 ("Mathematical probability is a system of reasoning that one must either use in its entirety or not use at all. There is no room for picking and choosing.").

as a result, the proposed likelihood approach to fact-finding sidesteps each of these bizarre results.

- Compared to the traditional probability model, close alternatives to the candidate hypotheses do not enter any part of the likelihood ratio; if the most likely pro-plaintiff theory is that the defendant was going 60 mph, then the existence of less likely theories (like 59 mph) simply has no bearing on the fact-finder's ultimate decision.
- Compared to the aggregate probability model, the likelihood theory involves a simple series of relative comparisons; in theory, it will always conclude with a single fully specified finding of fact.
- Compared to the probability ratio model, likelihood reasoning requires no modification of the standard properties of likelihoods; the yellow-light paradox does not apply because likelihoods do not sum to one (finding the evidence more consistent with a red light than a green light need not imply anything about the consistency of the evidence with a yellow-light hypothesis).

4.2 *Random Facts Problems*

A second type of problem for the probability approach arises from its conceptualization of *facts* as random variables. This implies that the facts of a case must obey the laws of probability, which leads to some strange and uncomfortable results.

A modest example of this type of problem is the property that continuous random variables have infinitesimal probability of assuming a given value. For example, the probability that a car was driving exactly 60 mph is technically zero: even if it was in that range, it might really have been 61 mph, 59.5 mph, or 60.001 mph—with enough evidence, any single guess can be almost surely disproved. This technical oddity complicates probability analysis, since it means that the probability of any combination of facts including a continuous variable will never exceed 50% (or even 0% for that matter). This problem is mitigated by aggregating probabilities of alternative factual theories, but that has its own problems, as discussed in the previous subsection.

An even more troubling random facts problem is what the literature has come to call the *conjunction paradox*. In brief, the conjunction paradox is the unsettling observation that when facts are treated as random variables, the threshold probability needed to prove individual facts diverges from the overall burden-of-persuasion for the case. A common example illustrates the problem.

Suppose a case with two disputed facts is being tried under the preponderance standard. To keep things simple, each fact is either true or false and the plaintiff wins only if both facts are true. At the close of evidence, the fact-finder concludes that the posterior probability of each fact being true is $P(f_1|e) = 0.7$ and $P(f_2|e) = 0.7$. Each fact is thus more probably true than false, but the probability of *both* facts being true—as needed for the plaintiff to win—may be much smaller than either individual probability. For example, if f_1 and f_2 are independent variables (so that neither fact suggest anything about the truth or falsity of the other), then the probability of both facts being true is $P(f_1 \ and \ f_2|e) = 0.7 \times 0.7 = 0.49$, meaning the plaintiff has actually failed to carry the burden of persuasion.

The conjunction paradox is a serious problem for any probability theory of fact-finding. First, it complicates description of the burden of persuasion, since the threshold probability for finding individual facts differs from that needed to decide the overall case, and actually depends on things (like the number of disputed issues) that will vary from one case to the next.[101] Second, it implies that the practical difficulty of satisfying a burden of persuasion should be strongly dependent upon the dubiously meaningful number of elements or material facts in a claim.[102] As an example, *theft* has more elements than *murder* in most jurisdictions, yet few see a viable argument for subjecting the

[101] *Cf.* Posner, *supra* note 14, at 1513–14 (oscillating on how a fact-finder would decide the case when individual and joint probabilities lead to different outcomes).
[102] *E.g.*, Sanchirico, *supra* note 12, at *19 ("Particularly troubling is the fact that the implied threshold probability for a charge, claim, or defense decreases (quite rapidly) in the number of elements it contains, a factor with uncertain relevance."); Allen, *supra* note 88, at 406–07 ("One implication of the conjunction principle is that it injects a certain inequality of treatment into the trial of disputes that is a function of the number of elements of a cause of action... plaintiffs' tasks will become more difficult as each new independent element is added.").

elements of theft to more a searching evidentiary standard.[103] Third, it suggests that the defendant's mere act of disputing an additional issue should have the surprising effect of increase the plaintiff's burden of persuasion on all *other* disputed issues as well.[104]

Few real attempts have been made to save the probability approach from the conjunction paradox. For example, it might be argued that the burden of persuasion should be applied to individual issues without regard to the probability of their joint truth.[105] Or that alternative factual theories should not be aggregated, so that the 0.49 probability of both facts being true can only be compared to the individual probabilities of just f_1, just f_2, or neither being true (0.21, 0.21, and 0.09, respectively).[106] But both of these arguments share the same deficiency: they amount to *ad hoc* modification of the rules of probability in a theory meant to derive from the rules of probability—a hint that probability may not be the right way to understand legal fact-finding.

Likelihood reasoning avoids the conjunction paradox. It does so by not treating the facts as random variables. Instead, likelihood reasoning treats the facts of the case as fixed, and the *evidence* as random. It then compares the probability of seeing the observed evidence under different factual hypotheses:

$$LR = \frac{L(f';e)}{L(f'';e)} = \frac{P(e|f')}{P(e|f'')}$$

[103] Leiter & Allen, *supra* note 2, at 1504–05.
[104] *See* Cheng, *supra* note 3, at 1263 ("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 [all other] elements at issue.").
[105] *E.g.* Alex Stein, *An Essay on Uncertainty and Fact-Finding in Civil Litigation, with Special Reference to Contract Cases*, 48 U. TORONTO L.J. 299, 311–12 n.27 (1998) (arguing that the conjunction paradox could be avoided by basing outcomes on the probabilities of elemental issues and ignoring their joint probability). Stein has since denounced this argument. *See* Allen & Stein, *supra* note 15, at 595–96.
[106] Cheng, *supra* note 3, at 1263–65. Cheng characterizes this solution as a result of comparing probability ratios, *id.* at 1263, but the actual argument around the conjunction paradox rests on an assumption that the defendant cannot aggregate probabilities of alternative facts to rebut the plaintiff's case, *id.* at 1264.

The product rule of probabilities still applies in the likelihood context, but it applies to the *evidence*, not the facts. And since the fact-finder is assessing the same evidence on top and bottom of the likelihood ratio, the product rule cancels itself out. Put another way, only the *relative* probability of the evidence under different factual theories matters, and the individual probabilities can be scaled up or down without any effect on their ratio. Intuitively, likelihood analysis treats the facts as conditioning parameters in probability statements: adding disputed facts changes the way the fact-finder thinks about the various factual hypotheses and their consistency with the evidence, but has no general effect on the ease or difficulty of meeting the burden of persuasion.

4.3 *Prior Probability Problems*

A third class of problems for probability theories of fact-finding arises from the failure of Bayesian probability analysis to discriminate between likelihood-relevant evidence and prior-probability information. Often illustrated by examples that suggest one result at an intuitive level, and another at a formal level, these problems appear to reflect an underappreciated sensitivity in the way that information is processed in legal fact-finding. Two popular examples are the well-known *Gatecrasher* and *Blue Bus* paradoxes.

The Gatecrasher paradox is a toy fact-pattern that leads to a strange result under traditional probability reasoning.[107] Suppose 1,000 people are in the stands of a rodeo, but a look in the cash register shows that only 499 of them have paid the price of admission. Nothing indicates who paid and who didn't, but it is clear that 501 people have jumped the gate. It has been noted—many times over—that the probability a randomly chosen attendee would be one that jumped is 50.1%, so that any randomly chosen attendee would be liable for the price of admission under the preponderance standard as applied in the probability

[107] *See* L. JONATHAN COHEN, THE PROBABLE AND THE PROVABLE 75 (1977) (giving the first description and illustration of this paradox).

approach. Everyone agrees that Bayesian logic compels this outcome, but no one seriously thinks it is the right result.[108]

The Blue Bus paradox is a slightly different situation, loosely based on the facts of an actual case.[109] The plaintiff is driving home at night when a reckless bus driver forces her off the road and into a ditch. The plaintiff only has time to note the color of the bus—blue. She sues the local Blue Bus Company, alleging that it owns 80% of all blue busses in the area. None of this is contested, and no other evidence is put forth by either side. Many commentators conclude that this means the Blue Bus Company has an 80% probability of responsibility.[110] But, again, few are comfortable assigning liability on this record alone.[111]

Efforts to defend the Bayesian probability approach against these paradoxes have been strained and unpersuasive. A common argument is that the fact-finder can escape the requisite liability conclusion by drawing a negative inference from the absence of better evidence in both puzzles.[112] But this is tantamount to assuming away the problem. Other arguments suggest that auxiliary policy objectives—like judicial economy or process validity—may require the plaintiff to show more

[108] *E.g.* Allen & Stein, *supra* note 15, at 573–74 ("[The attendee's] claim that he actually paid for his admission to the rodeo only has a 0.499 probability. Hence, under the preponderance standard … the organizers appear to be entitled to recover [from the randomly chosen attendant], which is patently absurd.").

[109] Smith v. Rapid Transit, Inc., 58 N.E.2d 754 (Mass. 1945). The hypothetical given in the text reflects Lawrence Tribe's stylized version of this case. Tribe, *supra* note 8, at 1341–42, n.7.

[110] The standard *Blue Bus* hypothetical actually provides insufficient information to compute an unambiguous probability of responsibility. *Cf. infra* note 115.

[111] *E.g.* Posner, *supra* note 14, at 1508–09 (interpreting a similar hypothetical in a way that implies the Blue Bus Company would have a posterior probability of liability of 80%, and noting the intuitive absurdity of this result, at least in cases where the posterior probability of liability is not too great).

[112] *E.g.* David Kaye, *The Paradox of the Gatecrasher and Other Stories*, 1979 ARIZ. ST. L.J. 101, 107–08 (1979) (arguing that the fact-finder may draw an inference from the lack of other information in the hypothetical); Posner, *supra* note 14, at 1509 ("The problem that causes this disbelief, however, is not with mathematical probability but with the tacit assumption that the statistic concerning the ownership of the buses is the only evidence that the plaintiff can obtain.").

than technically needed to win in these scenarios.[113] Perhaps, but this is again dodging the core difficulty of each puzzle. Instead, the clearest view of the situation is that the *paradox* in each puzzle comes from the effort to draw fact-finding conclusions from prior probabilities, and not from likelihood-relevant evidence.

This is easiest to see in the Gatecrasher paradox. Consider how the situation would be formalized under Bayes' Theorem for a randomly chosen rodeo attendee:

$$\frac{P(jumped \mid e)}{P(paid \mid e)} = \frac{P(e \mid jumped)}{P(e \mid paid)} \times \frac{P(jumped)}{P(paid)}$$

The posterior probability ratio of jumping to paying (left term) equals the likelihood ratio for the evidence (middle term), multiplied by the prior probability ratio of jumping to paying in the overall group of attendees (right term). The likelihood ratio is exactly one: the only "evidence" in the puzzle is the ticket box, and the contents of the box would look the same whether the defendant was one of the paying attendees or one of the gate jumpers. Rather, this is the rare case of a clear prior probability: there is a 50.1% chance that a randomly chosen attendee would be one that had jumped the gate. Thus, the posterior probability ratio equals a likelihood ratio of one, times a prior probability ratio of $50.1/49.9 = 1.004$, proving any randomly chosen attendee liable under the preponderance standard of the Bayesian probability theory of fact-finding.

Discomfort with that outcome belies likelihood reasoning. Liability is premised entirely on prior probabilities in the above analysis. But prior probabilities do not factor into likelihood analysis at all. With a likelihood ratio of exactly one, there is insufficient evidence (literally no evidence) to meet the preponderance standard of the likelihood theory of fact-finding. So, in contrast to the paradox that results from Bayesian probability reasoning, a randomly chosen attendee would not be found liable under a likelihood approach to the puzzle.

[113] *E.g.* Posner, *supra* note 14, at 1509 (arguing that judicial economy may require the plaintiff to expend more effort in these cases); *see also* Wells, *supra* note 42, at 740 (noting and citing various similar policy arguments).

This difference in outcomes reflects a fundamental difference in the way each approach handles information about average behaviors and background frequencies. As noted above, Bayesian probability analysis holds a randomly chosen attendee liable on the support of information about prior probabilities: the ticket box describes the average behavior of all the rodeo attendees. Likelihood analysis rejects any reliance on prior probabilities (averages or background frequencies) and instead focusses exclusively on information generated by the individuals and the transaction in dispute. This is perhaps best illustrated by considering two extreme variations of the standard Gatecrasher puzzle.

First, suppose that only one ticket is missing from the ticket box, but a ticket agent takes the stand to testify that she thinks she saw the defendant jump the gate. This testimony is not strong evidence of the defendant's liability, but if that's all there is—and if the testimony is given full credit by the fact-finder, and not, for example, assumed to be a self-serving lie—then it alone is sufficient to carry the preponderance standard under likelihood analysis. It makes no difference that only one attendee in the crowd jumped the gate: the inquiry is about *this* defendant, and the only evidence points toward liability.

Second, suppose there is no testimony to be had, but a look in the ticket box actually revealed one lonely ticket—the implication being that fully 999 of the 1,000 attendees had jumped the gate, so that the prior probability a randomly selected attendee would be a gate jumper is now 99.9%. As in the standard Gatecrasher puzzle, the situation here presents no *evidence* in the likelihood sense of the term. One attendee clearly paid the price of admission, and nothing has been introduced to show that this particular attendee is more likely a jumper than the one payer. So a likelihood approach would not find a randomly chosen attendee liable. Of course, social policy might suggest holding *all* members of the group jointly liable in such an extreme case.[114] But that only clarifies the point: available information proves *the group* liable, not any particular *individual* therein.

---

[114] *See, e.g.*, Summers v. Tice, 33. Cal.2d 80 (1948) (providing the classic example of alternative liability, in which proof that some member of a group caused an injury shifts the burden to the group members to prove their individual innocence).

Similar logic applies in the Blue Bus case. The cleanest way to frame this puzzle is as a prior probability ratio of 80%/20% in favor of the Blue Bus Company's responsibility, with a likelihood ratio of, again, exactly one.[115] As in the Gatecrasher puzzle, this lack of likelihood-relevant evidence prevents the fact-finder from holding the Company liable. But consider a related puzzle proposed by Gary Wells: the same circumstances as before, but instead of the Blue Bus Company owning 80% of the blue buses, a weigh-station logbook notes a Blue Bus Company bus passing down the road just before the collision, though the logbook is shown on cross examination to be accurate only 80% of the time.[116] The puzzle now contains likelihood-relevant evidence: the logbook has an 80% probability of correctly identifying the Blue Bus Company if the Company was responsible for the incident, and only a 20% probability of falsely identifying the Blue Bus Company if it was not responsible ($LR = 0.8/0.2 = 4$). Likelihood reasoning would thus assign liability in the logbook version of the puzzle ($LR = 4$), but not in the canonical version ($LR = 1$).

This difference in outcomes again reflects the distinction between prior probability and individualized evidence in likelihood reasoning. Information about background frequencies and averages (ownership rates for blue busses which would be true whether the Blue Bus Company was responsible or not) has no relevance in likelihood analysis; only individualized evidence (logbook information which would more probably be observed if the Blue Bus Company were responsible than if it were not) factors into the likelihood ratio. Thus, lack of individualized evidence saves the Blue Bus Company in a likelihood approach to the canonical version of the Blue Bus puzzle, and individualized evidence condemns it in the logbook version.

[115] Not enough information is provided to be certain how to frame the Blue Bus case. Depending on assumptions about the prudence of drivers and the frequency of other bus colors, scenarios consistent with the puzzle can range from strong proof of liability, to strong proof against liability. To illustrate, consider the puzzle-consistent situation where the Blue Bus Company has 8 blue busses and 990 red busses, while the only other bus company in the area has 2 blue busses and nothing else.
[116] Wells, *supra* note 42, at 741.

These examples illustrate the likelihood theory's sensitivity to the nature of information being processed: likelihood-relevant evidence is relied upon, while prior probability is not. Admittedly, the distinction between what is evidence and prior probability is not always clear. And whether something characterizes evidence or prior probability often depends on how fact-finding questions are framed.[117] But while drawing a sharp distinction on potentially so subtle a difference may seem like a defect of the likelihood theory of fact-finding, it is arguably one of its greater strengths.

First, whether ideal or not, the sharpness of this distinction reflects empirical reality. The whole reason that the Gatecrasher and Blue Bus puzzles are *paradoxes* in the first place is that the conclusions of Bayesian probability analysis are at odds with our intuition about how these cases should turn out. Applying likelihood reasoning to the puzzles yields conclusions better aligned with our expectations. This suggests that fact-finding may already involve likelihood reasoning—and it is not the only evidence to that effect.

In comparing the canonical- and logbook-version of the Blue Bus puzzle in a series of psychology experiments, Wells finds that subjects assess the same posterior probabilities in both versions of the puzzle, yet assign liability *frequently* in the logbook version, and *rarely* in the canonical version.[118] This is inconsistent with fact-finding based on Bayesian probability analysis: the subjects assess the same posterior probabilities in both versions of the puzzle, yet reach different results at the fact-finding stage. Rather, these results are consistent with sub-

---

[117] Suppose a negligence case arises from a traffic collision. The defendant argues that she was driving carefully when her brakes randomly failed. The plaintiff replies with information on the rarity of random break failure. One view of this showing is that it describes the prior probability of the defendant's theory. Another view is that the showing helps the fact-finder to contrast the likelihood of seeing the evidence (a traffic collusion) under the competing theories that the defendant was driving negligently (for which a crash is relatively probable) as opposed to driving prudently (for which a crash from random brake failure is relatively improbable).

[118] *Id.* at 742 (figure 1); *see also id.* at 744 (figure 3).

jects reasoning in likelihood terms—and thus demanding individualized, likelihood-relevant evidence to support a finding of liability.[119]

Second, the distinction between individualized, likelihood-relevant evidence and prior probability reflects existing applications of the law of evidence. As an example, consider the ban on character reasoning: evidence of a person's character is generally inadmissible to prove action in conformity with that character on a particular occasion.[120] Put another way, a person's character-defined prior probability of taking some action is irrelevant in deciding whether that person did so act in a given dispute. Evidence law thus recognizes and proscribes one form of prior-probability information already.

But an even clearer example is judicial treatment of *naked statistical evidence*. When faced with questions about the introduction of purely statistical evidence, many courts and fact-finders have declined to rely upon statistics that are not in some way individualized to the specific parties or transaction at issue.[121] An old state-court opinion colorfully captures the general tenor of this thinking:

> That in one throw of dice there is a quantitative probability, or greater chance, that a less number of spots than sixes will fall uppermost is no evidence whatever that in a given throw such was the actual result. … The slightest real evi-

[119] *Id.* at 746 ("The hypothesis offered here is that in order for evidence to have a significant impact on people's verdict preferences, one's hypothetical belief about the ultimate fact must affect one's belief about the evidence."); *id.* at 750 ("[I]t could be argued that people will allow their subjective probabilities to drive their verdict decisions only if the evidence on which those subjective probabilities are based is responsive to assumptions about the ultimate fact.").
[120] *E.g.* FED. R. EVID. 404.
[121] *E.g.* U.S. v. Shonubi, 103 F.3d 1085, 1092 (2d Cir. 1997) (distinguishing "specific evidence" of the defendant's conduct from background statistics on what "117 other people had done" under similar circumstances); *cf.* Allen, *supra* note 95, at 1099 ("If a statistic has no counterfactual implications, if it really is just an accidental property, then it tells us nothing about an event that is not in the particular set that generated the statistic.").

dence that sixes did in fact fall uppermost would outweigh all the probability otherwise.[122]

This insistence on individualized "evidence" is bewildering from a Bayesian probability perspective,[123] since background averages (prior probabilities) and individualized information (likelihood-relevant evidence) *both* factor into the computation of posterior probabilities. But the requirement of individualized evidence is easily explained by the likelihood theory of fact-finding. Indeed, efforts to define disfavored naked statistical evidence provide a useful rule of thumb for distinguishing prior probabilities from evidence: prior probabilities involve information that is "not case specific in the sense that the evidence was not created by the event in question but rather existed prior to or independently-of the particular case being tried."[124]

Third, to the extent that the distinction between likelihood-relevant evidence and prior probability reflects common fact-finding practice, it is important that a theory of evidence encompass this distinction. While it may often be intuitive what information constitutes likelihood-relevant evidence as opposed to a prior probability, some cases will inevitably fall close to the line. And in those cases, it is better that the legal community recognize and struggle to enforce the difference between prior probability and likelihood-relevant evidence, than for this distinction to be simply swept under the rug as it would be in a Bayesian probability approach to fact-finding.

4.4 *Personal Belief Problems*

The final problem with the probability approach to fact-finding is also

---

[122] Day v. Boston & M.R.R., 96 Me. 207 (1902).

[123] *E.g.* Jonathan J. Koehler, *The Normative Status of Base Rates at Trial*, *in* INDIVIDUAL AND GROUP DECISION MAKING: CURRENT ISSUES 137, 141 (N. J. Castellan, Jr. ed., 1993) (relegating the idea that prior probabilities are irrelevant to fact-finding, because they only describe group or long-run behavior, to the status of an argument "more likely to be advanced by law students or practicing attorneys who have little or no familiarity with statistics and probability theory").

[124] Wells, *supra* note 42, at 739 (providing this as a typical definition for "Naked statistical evidence," though noting that the term is ill defined in the legal literature").

arguably the most important: its reliance on prior probabilities makes the outcome of fact-finding dependent on the subjective prior beliefs and biases of the fact-finder. One way to see how deeply prior beliefs shape fact-finding is to transform the posterior probability ratio from Bayesian analysis into likelihood terms.[125] Suppose two factual theories are being compared under Bayesian probability reasoning: $f' \in C$ and $f'' \in C^c$. The fact-finder would find for the plaintiff if

$$PR = \frac{P(f'|e)}{P(f''|e)} = \frac{L(f';e)}{L(f'';e)} \times \frac{P(f')}{P(f'')} > m$$

This is simply Bayes' Theorem with the likelihood-ratio (middle term) expressed in likelihoods. The threshold, m, is the standard probability threshold for whatever burden of persuasion applies to the case.[126]

By rearranging terms, the above posterior-probability test can be transformed into a specific type of likelihood-ratio test:[127]

$$LR = \frac{L(f';e)}{L(f'';e)} > m \times \frac{P(f'')}{P(f')}$$

That is, a Bayesian probability theory of fact-finding is equivalent to a form of likelihood theory in which the evidentiary threshold needed to convict or assign liability varies with the fact-finder's prior beliefs. Specifically, it is a likelihood approach in which the burden of persuasion is defined as a fixed constant (the formal burden of persuasion) that gets scaled up (raising the burden) or down (lowering the burden) according to the fact-finders' personal prior beliefs.

The idea that the burden of persuasion would vary by fact-finder or context is troubling for at least two reasons. First, in contrast to some fields—where the influence of prior beliefs might soon be washed away by the collection of enough data—there is no reason to believe that prior beliefs will become so diluted by the evidence in the typical

---

[125] This presentation of posterior probability reasoning in terms of likelihood ratios is similar to that of Kaplow, *supra* note 49, at 34–35.
[126] *See supra* notes 39–40 and accompanying text.
[127] This likelihood ratio test comes from the part of the above probability ratio test to the right of the equality: simply divide by the prior probability ratio to get the result.

fact-finding exercise.[128] Second, and even more fundamentally, while there is nothing inherently debatable about how a fact-finder feels at the start of the case, the desirability of building such personal feelings into the fact-finding process is another matter altogether.[129] It seems fair to insist that the underlying epistemology of legal fact-finding should be a public concept, not a private one.[130] If this is too abstract, some examples may illustrate the underlying concern.

One example was already given in the introduction of this article. It concerned a hypothetical judge who started trial with a strong personal belief in the defendant's liability, and who thus found the defendant liable despite an overall evidentiary record that everyone would agree weighed in the defendant's favor. The judge's prior beliefs lowered the effective burden of persuasion so far that the balance of the evidence did not even need to favor the plaintiff for the judge to assign liability. This seems patently unfair, but it is actually rather tame in comparison to some of the more shocking implications that the reliance on prior beliefs entails for legal fact-finding.

One such implication is the procedural endorsement of fact-finder bias. The recent Supreme Court case of *Peña-Rodriguez v. Colorado* provides a timely example. This case involved a fact-finder (a juror) who explained during deliberations that the Latin-American defendant in the case could be presumed guilty of sexual harassment charges because, "in [this juror's] experience as an ex-law enforcement officer, Mexican men had a bravado that caused them to believe they could do

---

[128] *See* Leiter & Allen, *supra* note 2, at 1508 ("[I]ndividuals can begin from radically different perspectives, and each, in Bayesian terms, will be operating equally rationally. … In other contexts, such as science, these differences may be marginalized by convergence theorems that demonstrate that over time and with enough new evidence … result will [still] converge on the truth. There is nothing even remotely analogous to this in the condition of trials.").

[129] *See* Pawitan, *supra* note 19, at 13 ("There is nothing really debatable about how one feels … [but] one's formal action based on such feeling is open to genuine disagreement.").

[130] *Cf.* Taper & Lele, *supra* note 85, at 513 ("[Bayesian analysis] is held by many [to be] the most appropriate method of developing personal knowledge. This may be, but ... Science depends on a public epistemology not a private one.").

whatever they wanted with women."[131] The juror went on to share his belief that Mexican men were physically controlling of women, and to conclude that "I think he did it because he's Mexican and Mexican men take whatever they want."[132] Any doubt that this juror was using his personal belief in describing a prior probability of guilt is dispelled by the juror's further explanation that "in his experience, 'nine times out of ten Mexican men were guilty of being aggressive toward women and young girls.'"[133] Suppose, for sake of argument, that these (odious and antiquated) beliefs were sincerely held.

Any system of fact-finding based on the fact-finder's personal belief about the facts, necessarily starts from the fact-finder's prior beliefs, including those based on racial bias, gender stereotypes, assumptions about religious groups, etc. The Bayesian probability approach to fact-finding requires the use of these beliefs.[134] And in doing so, it condones the differential treatment of defendants. In *Peña-Rodriguez*, for example, the probability approach to fact-finding would essentially instruct this fact-finder to consider only the evidence presented at trial, but to demand a smaller quantum of evidence to convict a Mexican defendant than (presumably) a white defendant. I reject that even the most ardent proponent of Bayesian analysis would support this implication.

Nor do I believe that proponents of probability reasoning really do have any confidence in the use of prior beliefs in fact-finding.[135] The problem is that there is no way to avoid these implications and still retain a probability interpretation of the fact-finding process. Cheng,

---

[131] Pena-Rodriguez v. Colorado, 137 S. Ct. 855, 862 (2017).
[132] *Id.*
[133] *Id.*
[134] *See* Cheng, *supra* note 3, at 1267 ("In legal venues, one might fear that [prior probability ratios] embody prejudices against certain types of parties."); *cf.* Posner, *supra* note 14, at 1514 ("Ideally we want the trier of fact to work from prior odds of 1 to 1 that the plaintiff or prosecutor has a meritorious case. A substantial departure from this position, in either direction, marks the trier of fact as biased.").
[135] *But cf.* Posner, *supra* note 14, at 1494–95 ("[If] the judge's prior odds are 100 to 1 in favor of guilt, [and] the evidence creates a likelihood ratio of 8 to 1 that the defendant is not guilty, [then] the judge's posterior odds on guilt will still be 12.5 to 1. All this is perfectly rational.").

for example, has attempted to argue that prior probability ratios are always equal to one in a fact-finding setting: "the legal system imposes a constraint … [that normatively fixes] the prior odds ratio at l to start the plaintiff and the defendant in equipoise."[136] Does this make sense? Cheng's justifications for the claim are that it equates to unbiased fact-finding,[137] is (somehow) implied by the plaintiff's satisfaction of the burden of production,[138] and is what fact-finders ought to do, even if they don't do it.[139] These arguments are not inspiring. But even if they were, the proposed reliance on uninformative or uniform prior probabilities would still not be the simple fix that it appears to be.

First—at a practical level—the argument confuses what society gets to control in the fact-finding setting. Fact-finders are not robots that can be programed to have arbitrary personal beliefs. At least for now, all fact-finders are human. And while efforts at persuasion may be effective in some circumstances, in general we cannot ask a person for their *personal belief* after having seen the evidence, without the result being informed by their personal belief before seeing the evidence. If the juror in *Peña-Rodriguez* sincerely believed the things he said, then these priors will necessarily inform his posterior beliefs. This is not a claim that the juror would attempt to subvert the legal process to act on his bias; it is simply a truism that he has no prior beliefs to use except his own when assessing his posterior beliefs about the facts.

Second—at a technical level—uniform prior probabilities are not the reflection of initial-state ignorance that they appear to be. To say "I do not have any *a priori* idea whether the defendant is a murderer" is

---

[136] Cheng, *supra* note 3, at 1273.
[137] *Id.* at 1267 n.24 (citing Posner, *supra* note 14, at 1514).
[138] Cheng, *supra* note 3, at 1267–68 ("As long as the plaintiff articulates a prima facie case and satisfies the burden of production, the case starts with both parties in equipoise."). I fail to grasp the logic of this claim, but even if it were true, it would merely push the problem of the fact-finder's prior beliefs to the stage of determining whether the plaintiff has met the burden of production.
[139] *Id.* at 1267 n.24. As a normative basis for this claim, Cheng and Pardo suggest it would minimize the rate of fact-finding errors under certain distributional assumptions about the evidence-generating process. Edward K. Cheng & Michael S. Pardo, *Accuracy, Optimality and the Preponderance Standard*, 14 LAW, PROBABILITY & RISK 193 (2015).

not the same as saying "I think the defendant has an equal probability of being a murderer or a law-abiding citizen."[140] To illustrate that point, suppose the defendant is on trial for two independent charges of murder, and I know nothing about either charge. Assigning uniform prior probabilities to each individual charge means that I must necessarily assess a prior probability of 75% that the defendant committed *at least* one of the two murders—so that apparent ignorance in one respect becomes gratuitous information in another.[141]

The likelihood approach to fact-finding again escapes the problems of the Bayesian probability approach. First, likelihood reasoning has nothing to do with personal prior beliefs; it is reasoning from evidence alone. It is the evidence-theory answer to a question posed by Justice Sotomayor during oral arguments for *Peña-Rodriguez*: "Don't we want deliberations on evidence and not deliberations on someone's stereotypes and feelings about the race of a defendant?"[142] Of course we do. Likelihood reasoning is precisely this approach to fact-finding.

Second, the likelihood theory of fact-finding avoids the unsettling notion that different defendants may face different effective burdens of persuasion. In the likelihood approach to fact-finding, the fact-finder always compares the same evidence to the same threshold, regardless of the personal feelings and convictions of the fact-finder. Uniform prior probabilities are just another respect in which the literature has been grasping for something other than Bayesian analysis as the basis for a theory of fact-finding. Likelihood is what it's looking for.

5 CONCLUSION

This article stakes the immodest claim that more than a half century of

[140] *See* Royall, *supra* note 20, at 174 ("The reason why pure ignorance cannot be represented by a probability distribution is that every probability distribution represents a particular state of uncertain knowledge; none represents the absence of knowledge. … It is one thing to state that I do not know which of two possible values of θ is true, or that I have no knowledge or no prior evidence about which is true. It is quite another to assert that the two values are equally probable. ").
[141] *See* Edwards, *supra* note 16, at 58.
[142] Transcript of Oral Argument at 40:5–7, Pena-Rodriguez v. Colorado, 137 S. Ct. 855 (2017) (No. 15-606).

research on fact-finding and the burdens of persuasion has focused on the wrong thing. It seems that most fact-finding is not concerned with Bayesian probabilities at all. It is concerned with likelihood reasoning. The objective of this article is to propose a new likelihood theory of fact-finding, and to thus get the conversation back on track.

This is not to say that this likelihood theory of fact-finding is without limitations. Like any formal model, it is an abstraction that fails to encompass much of the nuance and complexity of trial fact-finding. Likelihood reasoning is also limited in that it works from within a frame of reference, but cannot itself guide the framing of a question. In practice, this means that what is or isn't prior probability information will always be susceptible to argument and difficult cases, and in these cases the concept of likelihood will not itself resolve disputes. Finally, this likelihood theory is an awkward fit to the specific circumstance of prospective fact-finding—the prediction of future outcomes from historic evidence. In this specific situation, there may yet be room for prior belief and Bayesian probability analysis in fact-finding.

But these limitations are narrow in scope, and the benefits of the proposed likelihood approach are meaningful. The proposed theory of fact-finding encompasses every burden of persuasion in use today, reflects the realities of adversarial fact-finding, supports and extends empirical models of fact-finders' cognitive processes, avoids the problems and paradoxes of the Bayesian probability approach, and affords fresh insights into the fact-finding process and the effects of subjective personal beliefs and biases when reasoning under uncertainty.

So what are the immediate practical implications of this new way of understanding legal fact-finding? To be clear: I am not at all proposing that fact-finders should be instructed in the rigors of likelihood analysis before entering the courtroom. Nor am I proposing any semblance of the feared *trial by numbers* in which calculators replace pads and pencils.[143] I also do not think that simply changing the language from *probability* to *likelihood* will change anything in practice. The difference between these concepts is too foreign and too esoteric for mere words to make any difference on their own.

[143] Calculators *with* pads and pencils, I support.

In common with all work on the theory of fact-finding, and born of a conviction that "Whatever enables lawyers to think more clearly is of practical importance,"[144] the objective of this article is to bring clarity and understanding to the fact-finding process. Clarity does not always call for reform, and to the extent that empirical models of fact-finders' cognitive process are correct, fact-finders may already be reasoning as this likelihood theory would hope in many cases. Accidental success is weak assurance against mistake, though, and the proposed likelihood theory *does* imply serious changes in how lawyers, judges, and scholars need to understand, apply, and describe the fact-finding process, and the conventional burdens of persuasion.

And while changing mere words will make little difference, the legal community must *understand* the distinction between probability and likelihood. Probability terms are now scattered throughout the law of evidence. The federal rules of evidence define "relevant evidence" as anything having the "tendency to make a fact [of consequence] more or less probable than it would be without the evidence."[145] The model rules of evidence define "finding a fact" as "determining that [the fact's] existence is more probable than its non-existence."[146] And some scholars go so far as to claim that "a lawsuit is essentially a search for probabilities."[147] The language being used is irrelevant, but the concept is important. Bayesian probability is not the measure of these ideas. Relevance, fact-finding, and lawsuits all involve the weighing of evidence in order to assess the relative plausibility of competing theories about the world. Likelihood is the measure of this endeavor.

That distinction is not a mere technicality. Bayesian probability is fundamentally a description of belief; likelihood is fundamentally a description of evidence.[148] Richard Royall provides this helpful way to conceptualize the difference:[149]

---

[144] Lempert, *supra* note 13, at 1056.
[145] Federal Rule of Evidence 401.
[146] American Law Institute, Model Code of Evidence, Rule 1(5) (internal quotation marks omitted).
[147] McCormick on Evidence, *supra* note 36, at 669.
[148] Royall, *supra* note 9 at 131.
[149] *Id.* at 122 (describing and annotating similar questions).

- Posterior probability answers the question: "What do I believe, now that I have seen this evidence?"
- Likelihood answers the more basic question: "What does this evidence show?"

Over the years, many scholars have written that fact-finders' beliefs are the thing we care about.[150] But the work of this article has been to show the exact opposite. Not only do we not care about fact-finders' beliefs, the role of prior beliefs should be minimized and excluded from the fact-finding process as far as it possibly can be. Contrary to decades of legal thinking, the only way to make sense of the fact-finding process is to focus exclusively on weight-of-evidence alone.

Finally, the proposed likelihood theory of fact-finding suggests the need for changes in the way that burdens of persuasion are described. Again, simple word changes probably won't matter much, but the examples, illustrations, and elaborations meant to clarify these burdens need revision. While fact-finders should, as ever, use their experience and intuition to guide their thinking, the likelihood approach does not ask for their personal beliefs about what happened. The fact-finding inquiry is limited to their assessment of the evidence produced at trial. Their task is not to gauge their personal beliefs about the facts, but to instead compare the relative consistency of the observed evidence with different theories about the facts in dispute. At root—and particularly when operating under the preponderance-of-the-evidence standard—this task is a guided search for the most likely story.

[150] *E.g.* James, *supra* note 94, at 53 ("All would agree that what counts is the jury's belief in the existence (or non-existence) of the disputed fact, and the extent to which the evidence actually produces that belief; surely we are not seeking the jury's estimate of the weight of evidence in the abstract…"); McBaine, *supra* note 1, at 247 ("The [common jury instruction speaks of] the weight of the evidence. It does not, as it should do, direct [attention] to the degree of belief which the proponent of the proposition must produce [before he is] entitled to a finding favorable to him.").