

A LIKELIHOOD STORY

THE THEORY OF LEGAL FACT-FINDING¹

Sean P. Sullivan²

February 1, 2018

Are racial stereotypes a proper ground for legal fact-finding? If not, what about gender stereotypes, sincerely believed by the fact-finder to be true, and informed by the fact-finder's own life experience? What about statistical evidence of culpability? If people of the defendant's gender, education level, job title, and past criminal history exhibit a statistically greater incidence of violent behavior than the population overall, is this evidence that the defendant acted violently on a particular occasion?

The intuitive answer is that none of these inferences feel like proper bases on which fact-finders should be deciding cases. But why not? Nothing in the traditional probability or belief-based theory of fact-finding justifies any such exclusion. Maybe intuition goes astray here. Or perhaps the traditional theory of fact-finding is wrong. Arguing the latter, this article proposes a new theory of fact-finding. Rather than being based on probability or beliefs, this theory suggests that idealized fact-finding is an application of likelihood reasoning—the statistical analog of the old legal concept of the “weight of evidence” in a case, and the theoretical analog of modern descriptions of fact-finding as a process of assessing the relative plausibility of competing factual stories on the observed evidence.

¹ This paper has benefitted from the thoughtful comments of Ron Allen, Ed Cheng, Kevin Clermont, Ken Elzinga, Carrie Figdor, Josh Fischman, Brandon Garrett, Mike Gilbert, Josh Goodman, Dan Greenfield, Jay Koehler, Greg Mitchel, Michael Pardo, John Pepper, Devesh Raval, Paul Rothstein, Steve Salop, Bob Tovsky, and workshop participants at the Federal Trade Commission, the University of Iowa, the University of Virginia, and the 2017 meeting of the American Law and Economics Association.

² University of Iowa College of Law. Contact the author at sean-sullivan@uiowa.edu.

This likelihood theory of fact-finding marks a fundamental change in the way legal fact-finding should be understood, and from this change come a number of advancements. This likelihood theory of fact-finding simplifies the concept of fact-finding, describing every burden of persuasion as an application of the exact same reasoning principle. It harmonizes recent scholarship on fact-finding, showing how empirical work on the cognitive processes of fact-finders can be formalized into an intuitive and coherent theory of the ideal fact-finding process. It explains the existence of fundamental rules of evidence, like the inadmissibility of character evidence. It explains traditional mores of fact-finding, like hostility to naked statistical evidence. And, most importantly, it affords new insights into the effects of subjective beliefs on fact-finding, clarifying and offering a solution to the problem of fact-finder bias in particular.

Table of Contents

1	INTRODUCTION	3
2	THE LANDSCAPE OF UNCERTAINTY AND PERSUASION. 10	
	2.1 Probability: Absolute, Subjective Belief.....	10
	2.2 Likelihood: Relative Weight of the Evidence	13
	2.3 Persuasion: Fact-Finding under Uncertainty.....	16
3	A LIKELIHOOD THEORY OF FACT-FINDING	20
	3.1 The General Fact-Finding Framework	20
	3.2 The Likelihood Approach to Fact-Finding	24
	3.3 Intuition for the Likelihood Approach.....	32
4	COMPARISONS AND IMPLICATIONS	35
	4.1 Implications for the Ideal Fact-Finding Process	35
	4.2 Implications for the Conjunction Paradox	39
	4.3 Prior Probability, Character Reasoning, and Naked Statistics.....	42
	4.4 Personal Beliefs, Prejudice, and Biased Deliberations.....	51
5	CONCLUSION	56

1 INTRODUCTION

In oral arguments for the recent Supreme Court case of *Peña Rodriguez v. Colorado*, Justice Sotomayor voiced a rhetorical, but serious, question: “Don’t we want deliberations on evidence and not deliberations on someone’s stereotypes and feelings about the race of a defendant?”³ The answer is obviously “Yes!” But why was the question even asked? How could it be that over 200 years into the history of our country, a justice of the highest court of the land would have the occasion to ask a question with so startlingly obvious an answer?

Other questions only raised the level of alarm. Justice Alito asked whether juror discrimination on the basis of a party’s religion differed from racial discrimination.⁴ The Chief Justice pushed this point: “isn’t it odious to have [jurors] discriminat[e] against someone because he’s a Muslim or practices Islamic faith? ... [Saying,] I know how Muslims behave; he committed this crime.”⁵ Justice Ginsburg asked whether gender discrimination in an automotive case would be different: “Suppose [a juror says], what do you expect of women drivers? Women shouldn’t be allowed to drive cars. Every woman I know is a terrible driver.”⁶ Again, these questions shock only because their answers are so obvious. None of these inferences is different from another in substance; all are offensive to our sense of justice.⁷ To infer, for example, that because the defendant is a woman, she must be a bad driver—and thus to subject women defendants to what is effectively a lower burden of proof than their otherwise identical male counterparts in automotive tort claims—is so offensive to our concept of fairness as to be simply laughable.

³ Transcript of Oral Argument at 40:5–7, *Peña-Rodriguez v. Colorado*, 137 S. Ct. 855 (2017) (No. 15-606).

⁴ *Id.* at 5:10–14.

⁵ *Id.* at 6:16–22.

⁶ *Id.* at 12:1–6.

⁷ In deciding this case, the Court *distinguished* race-based discrimination from other forms of prejudice on grounds such as the history of race relations and race-specific jurisprudence in this country. *Peña-Rodriguez v. Colorado*, 137 S. Ct. 855, 868–69 (2017). The special harm of racial discrimination conceded, the tact of distinguishing racial bias from other forms of prejudice only defers the fundamental question: when, if ever, should prior prejudice be *permitted* in fact-finding—and why?

But, again, why were these questions even asked? The case involved an effort to impeach a jury's verdict, and the issue facing the Court was whether jurors should be allowed to testify about racially biased deliberation statements,⁸ in spite of the usual prohibition on juror testimony about deliberations in inquiries into the validity of the verdict.⁹ But the most salient feature of the oral arguments was the close availability of so many cringingly realistic examples of biased fact-finding. And while each of the justices appeared to agree that these examples involved unfairness, looming uncomfortably over the proceeding was a deeper and unanswered question: *why* are these inferences unfair?

Uncertainty over this question may go a long way in explaining the doubtless prevalence of biased fact-finding today. Jurors are typically instructed to use common sense and their own personal experience in their deliberations.¹⁰ But where personal experience gives way to prejudice is never explained.¹¹ Lack of clarity on this point seems to reflect the deep uncertainty that our legal system has long held about the entire fact-finding process.¹² A problem not helped—and actually made far worse—by growing consensus that fact-finding involves some form of probabilistic reasoning about the truth of historic facts.

⁸ See generally Pena-Rodriguez, 137 S. Ct. at 862.

⁹ See, e.g., FED. R. EVID. 606(b)(1) (providing, subject to limited exceptions, that in “an inquiry into the validity of a verdict or indictment, a juror may not testify about any statement made or incident that occurred during the jury’s deliberations.”). Many state rules of evidence include similar prohibitions.

¹⁰ See, e.g., Third Circuit, Model Civil Jury Instructions § 1.5 (2017), (“You should use your common sense in weighing the evidence. Consider it in light of your everyday experience with people and events, and give it whatever weight you believe it deserves.”); Third Circuit, Model Criminal Jury Instructions § 1.08 (2012), http://www.ca3.uscourts.gov/sites/ca3/files/2012%20Chapter%201_0.pdf (same); Seventh Circuit, Federal Civil Jury Instructions of the Seventh Circuit § 1.11 (2017), http://www.ca7.uscourts.gov/pattern-jury-instructions/7th_cir_civil_instructions.pdf (“You should use common sense in weighing the evidence and consider the evidence in light of your own observations in life.”); Seventh Circuit, Pattern Criminal Jury Instructions of the Seventh Circuit § 2.02 (2012), http://www.ca7.uscourts.gov/pattern-jury-instructions/7th_criminal_jury_instr.pdf (same).

¹¹ Cf. *United States v. Jones*, 808 F.2d 561, 568 (7th Cir. 1986) (concluding, without analysis or citation, that instructions to use common sense in deliberations “[do] not... invite a jury member to consider the evidence in light of personally-held... stereotypes or prejudices.”).

¹² See generally J.P. McBaine, *Burden of Proof: Degrees of Belief*, 32 CALIF. L. REV. 242 (1944) (noting widespread uncertainty about the theory of legal fact-finding).

To see the surprising connection between hateful bias and abstract probability, consider the following puzzle. Suppose 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. Now consider two simple questions. 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?

The answer to the first question is easy: the probability of drawing a white marble from the white cup is $5/6$ (about 83%). The answer to the second question is a harder. Most people would comfortably guess that I chose the white cup in the second question, but the exact probability of that choice is not obvious. While some people recall that there is a rule of probability for this calculation (Bayes' Theorem), few have an intuitive grasp of exactly how it works.

But, as you probably guessed, the situation is not quite that simple. There was indeed a $5/6$ probability of seeing a white marble on a future draw in the first question. But in the second question, it does not make much sense to talk about the *probability* of my choice at all. For one thing, this not a future event with value still to be realized: I've already chosen the cup. That historic choice is now a fact in the world, not a random variable, and the probability that I chose the white cup is now—in a literal sense—either zero or one. Moreover, even if you *did* persist in trying to compute the probability of my choice using the relevant rule of probability (Bayes' Theorem), you would have quickly found that you need to know something I did not tell you: my latent or *prior* probability of choosing the white cup over the red cup. But I gave no prior probability for this choice because it, too, was not random. I chose a cup by conscious, willful act. There was nothing probabilistic about it.

Undeterred, you might think that the way to solve the difficulty of the second question is to simply *assume* a probability distribution for my choice of cup. You might guess, for example, that there was a 50% chance that I would choose the white cup, and with this assumption you could complete the Bayesian computation. But recall that the context

for the puzzle is the growing consensus that legal fact-finding can be understood and explained in probabilistic terms. How would this translate to the legal context? If the facts at issue are historic acts and conscious decisions—which, like my choice of cup, are not governed by meaningful prior probabilities—then where do you get your prior probabilities from? What is the prior probability that someone committed a crime, or failed to exercise due care behind the wheel? Perhaps you would leave these questions to the fact-finder, to be supplied by personal experience and assumption. But you already see the problem. What is the prior probability that a Muslim man committed a crime? What is the prior probability that a woman was driving negligently at the time of the collision? It is no stretch to say that a *probabilistic* or *belief based* approach to fact-finding not only invites bias into the decision-making process,¹³ but arguably demands it.

This seems like reason enough not to use probabilistic reasoning as a model of legal fact-finding, yet courts and scholars have spent the last fifty years or more struggling to do just that: to describe and apply burdens of persuasion and fact-finding in terms of *probabilistic belief* that the defendant committed a crime, acted negligently, or breached a contract. Why? Part of the answer is certainly that, even today, we still lack a formal theory of fact-finding, and this is simply the most recent effort to supply that framework. But while we do need a theory of fact-finding, it hardly follows that probabilistic belief concepts are the answer. Even if unbiased, actuarially accurate, probabilities *could* somehow be calculated for the long-run chance that a person would commit a crime, why would this matter? Responsibility for conscious actions and choices is basic to our system of justice. We do not punish a murderer for having some demonstrable probabilistic tendency toward violence, but for the act of raising and lowering the knife on a particular occasion.

That is, a probabilistic approach to fact-finding not only invites bias into deliberations, but does so in an effort to advance the logically and

¹³ As explained later in this article, one way to interpret Bayesian probability analysis is as an idealized model of subjective beliefs. See *infra* Section 2.1 (explaining the close connection between Bayesian posterior probabilities and subjective belief concepts). Other interpretations of beliefs are also possible. See, e.g., Kevin M. Clermont, *Death of Paradox: The Killer Logic Beneath the Standards of Proof*, 88 NOTRE DAME L. REV. 1061 (2013) (describing fuzzy belief functions). Careful distinction between the various possible definitions of belief is not critical to the current argument.

equitably questionable goal of seeking to find historic *facts* by the fact-finder's belief about the *probability* of their truth—and problems with the approach go deeper still. Decades of efforts to build a Bayesian theory of fact-finding have suffered despairing setbacks: artificial dependence of outcomes 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 fact-finding exercises, to name just a few.¹⁴

These issues are well documented, and while some researchers still seek a probability theory to fact-finding,¹⁵ many have abandoned the effort. Forfeiting, at least temporarily, any search for a formal theory of fact-finding, many scholars have turned their focus to understanding fact-finders' cognitive process at an empirical and descriptive level.¹⁶ Exciting work on relative plausibility,¹⁷ narrative coherency,¹⁸ and story-based¹⁹ models of deliberation advances our understanding of how fact-finders think (iteratively constructing and comparing stories that explain and relate to the evidence), but has relatively little to say about the interpretation of higher burdens of persuasion, or the normative scope of idealized fact-finding and boundaries of unfair prejudice, and thus is not a substitute for a formal theory of fact-finding. Nor need we give up the search for a formal theory just yet. The problems with probability-based models of fact-finding is not that mathematics and probability theory are inherently irreconcilable with fact-finding.²⁰ Instead, I argue that the problem is that most fact-finding appears not to

¹⁴ 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).

¹⁵ See, e.g., Edward K. Cheng, *Reconceptualizing the Burden of Proof*, 122 YALE L.J. 1254, 1256 (2013).

¹⁶ See generally Michael S. Pardo, *The Nature and Purpose of Evidence Theory*, 66 VAND. L. REV. 547 (2013) (summarizing recent developments in evidence theory).

¹⁷ E.g., Ronald J. Allen, *Factual Ambiguity and a Theory of Evidence*, 88 NW. U. L. REV. 604 (1994).

¹⁸ E.g., Dan Simon, *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, 71 U. CHI. L. REV. 511 (2004).

¹⁹ 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).

²⁰ Cf. Lawrence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971).

reflect probabilistic reasoning at all. It appears to reflect reasoning in terms of a related, but different, concept: *likelihood*.

This article presents a likelihood theory of fact-finding. The change from a probabilistic approach is more than semantic. Richard Royall provides a helpful way to conceptualize the difference:²¹

- Bayesian 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?”

Another way to think about the difference is in terms of 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 simpler logic that the white cup is the more *likely* cup to have produced the evidence they’ve seen. This is likelihood reasoning!

I argue that most of legal fact-finding (efforts to decide the truth of historic facts and actions) involves likelihood reasoning.²² This is not the first paper to suggest the use of likelihood analysis in legal fact-finding,²³ but it *is* the first paper to describe all current fact-finding, and every one of the common burdens of persuasion, in terms of likelihood reasoning alone. To be clear, though, my argument is not that this or any formal model of fact-finding must be imposed on fact-finders in a trial setting. Nor do I argue that this likelihood theory of fact-finding captures every nuance of the legal process; like any formal model, the theory is an abstraction with important limitations.

Instead, what I emphatically argue is that a reasonable and internally coherent theory of fact-finding should guide our legal practice and procedure. The proposed likelihood theory of fact-finding serves that role. It improves upon efforts to explain fact-finding in terms of probabilities

²¹ See Richard Royall, *The Likelihood Paradigm for Statistical Evidence*, in *The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations* 119, 122 (Mark L. Taper & Subhash R. Lele eds 2004) (describing and annotating similar questions).

²² See H. Richard Uviller, *Evidence of Character to Prove Conduct: Illusion, Illogic, and Injustice in the Courtroom*, 130 U. PA. L. REV. 845, 845 (1982) (“The process of litigation is designed for the reconstruction of an event that occurred in the recent past.”).

²³ E.g., Louis Kaplow, *Burden of Proof*, 121 YALE L.J. 738 (2012).

and beliefs, 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 scholarship, showing how empirical work on the cognitive processes of fact-finders can be formalized into an intuitive and coherent theory. It explains rules of evidence like the exclusion of most character reasoning, and explains judicial hostility to naked statistical evidence. And, perhaps most importantly, it affords new insights into the problem of subjective fact-finder belief, clarifying the problem of fact-finder bias in particular.

The arguments, here, are formal and theoretical, but the implications are immediate and practical. We need to adjust the way we think about and describe legal fact-finding. The formal arguments for a likelihood theory of fact-finding are developed in Sections 2 and 3. Readers who would prefer to avoid this more technical exposition can skip to Section 3.3 for an intuitive summary of the framework. Sections 4 and 5 then trace some preliminary implications of the proposed likelihood theory of fact-finding. These are *preliminary* implications of a new theory on which more study is needed, and more is likely to be learned.

To both foreshadow and clarify what is at stake, here, consider one last puzzle, involving a claim of unjust enrichment tried without a jury. Before any evidence is presented, the judge admits that he's had business dealings with people from the defendant's country before, and on that experience, he is already 80% certain that the plaintiff's complaint is true. At trial, the plaintiff produces only weak evidence to support her claim, while the defendant puts forth strong evidence that she was not enriched, much less unjustly so. Everyone agrees that the *weight of the evidence* favors the defendant. Yet because his prior *belief* was so strong, the judge still personally *believes* the plaintiff's claim to be true, and thus finds for the plaintiff. This is a textbook example of probability or belief-based fact-finding. It is functionally equivalent to the biased deliberation at issue in *Peña-Rodriguez*. But is this really how fact-finding should work? If the process and its outcome strike you as unfair, unjust, and wrong, the likelihood theory of fact-finding helps explain why.

2 THE LANDSCAPE OF UNCERTAINTY AND PERSUASION

It is impossible to propose a likelihood theory of fact-finding without first saying something about the basic differences between probability and likelihood as concepts of uncertainty. This material is unavoidably formal, but the emphasis is on intuition and legal applications.

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:²⁴

$$\begin{aligned} P(\text{draw white marble} \mid \text{white cup chosen}) &= 5/6 \\ P(\text{draw white marble} \mid \text{red cup chosen}) &= 1/6 \end{aligned}$$

Decades of work on fact-finding have focused on trying to swap the order of variables and conditions in this type of probability statement. The fact-finder is assumed to know, at least subjectively, the probability of seeing the evidence given the facts. 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:

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

Instead, reversing the order of terms requires appeal to a fundamental law of probability: Bayes' Theorem.²⁵

²⁴ 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 .”

²⁵ 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).

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:²⁶

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

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 relative probabilities after seeing the evidence, equal the relative probabilities before seeing the evidence, times the likelihood ratio of events on the evidence (more on that later).

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 the last 50 years.²⁷ But while much of Bayes' Theorem has been tirelessly explained and re-explained in prior work,²⁸ one aspect of the model that always seems to get short shrift is the way that interpretation depends on the source of prior probabilities.

Things are straightforward when objective prior probabilities are available. As an artificial example, suppose we were to repeat the cup-choice puzzle, but this time I based my choices on a die roll: I choose the white cup on a roll of 1–4; the red cup on 5–6. This randomization provides a clear prior distribution for my choice: there is a 2/3 prior probability that I will choose the white cup. Upon randomly drawing a

²⁶ See, e.g., Chris William Sanchirico, *Evidence: Theoretical Models*, in X ENCYCLOPEDIA OF LAW AND ECONOMICS *7–9 (2012).

²⁷ 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).

²⁸ See, e.g., Lempert, *supra* note 27, at 1022–25; Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1486–87 (1999).

white marble, you could use Bayes' Theorem to calculate the relative probability that I had chosen the white cup in this example:

$$\frac{P(\text{white cup chosen} \mid \text{draw white marble})}{P(\text{red cup chosen} \mid \text{draw white marble})} = \frac{5/6}{1/6} \times \frac{2/3}{1/3} = 10$$

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 10 times as often as the red cup. This is not so much a statement about my choice in a 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 the objective, long-run, or average probabilities of events.

But what happens when clear prior probabilities are not available? In the actual cup-choice puzzle, I did not flip a coin, but consciously chose a cup. What prior probabilities describe this non-random choice? And 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 prior probabilities, but there is little hope of finding objective, long-run, or average probabilities of these events.²⁹

When clear and objective prior probabilities are unavailable, Bayesian analysis falls back on subjective prior probabilities:³⁰ probabilities that reflect personal assumptions or individual beliefs of the user. The interpretation 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. Combining the inputs of a person's initial beliefs and the observed evidence, Bayes' Theorem describes what a fully rational person—who happened to have these prior beliefs—should now believe upon seeing the evidence. And since prior beliefs vary from person-to-

²⁹ 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.”).

³⁰ 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.”).

person and context-to-context, the same evidence can produce starkly different posterior probabilities (that is, posterior beliefs) under applications of Bayesian probability analysis.

2.2 Likelihood: Relative Weight of the Evidence

Though the terms are synonyms in informal English,³¹ 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 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.³² 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 observation of a white marble drawn at random from the unidentified cup.

³¹ *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).

³² 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, $c > 0$.

Likelihoods are related to probability statements by definition. But likelihoods are not probabilities.³³ At a deep level, probabilities and likelihoods behave differently.

- Probabilities have individual meaning (a $\frac{1}{2}$ 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.³⁴
- 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.³⁵
- 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.³⁶

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.³⁷ But likelihoods have their own attractive properties, owing to the special features of likelihood ratios. Chief among these is the *Law of Likelihood*, the takeaway of a set of related arguments that, for a given probability model, seeing 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 :

³³ 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).

³⁴ 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 32.

³⁵ E.g., IAN HACKING, LOGIC OF STATISTICAL INFERENCE 50 (2016) (noting that likelihoods do not obey the Kolmogoroff axioms); Fisher, *supra* note 31, at 327 (similar).

³⁶ E.g., Royall, *supra* note 34, at 16–20 (explaining and detailing this property).

³⁷ See, e.g., Pawitan, *supra* note 33, 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)).

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

The Law further provides that the total strength of evidential support is reflected in the size of this likelihood ratio.³⁸

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:

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

A likelihood-ratio of 5 means that the available evidence is five times more consistent with my having chosen the white-cup than the red-cup. This ratio would have the same interpretation in a medical context, a physical science context, or a legal fact-finding context.³⁹ Likelihood ratios provide a general language for describing 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 for which one need not assume *anything* about prior probabilities. Likelihood analysis of the cup-choice puzzle thus imposed no restriction at all on my choice of cup. The analysis would work just as well even if *no prior probabilities* could describe my choice process.⁴⁰ In this way, likelihood reasoning differs in a fundamental way from probability reasoning: likelihood analysis has no need for subjective prior beliefs, and the output is likewise not a statement of belief. Likelihood concerns, and describes, the comparative consistency of different hypotheses with the evidence. On

³⁸ This description of the *Law of likelihood* is based on Ian Hacking's formulation. Hacking, *supra* note 35, at 48–66; *see also* Edwards, *supra* note 30, at 28–31 (similar); Royall, *supra* note 34, at 1–3 (similar). For a formal argument of the related likelihood principle, see Allan Birnbaum, *On the Foundations of Statistical Inference* 57 J. AM. STAT. ASS'N 269 (1962).

³⁹ *See* Royall, *supra* note 34, at 27.

⁴⁰ *See* Edwards, *supra* note 30, at 55–65 (discussing the impossibility of representing ignorance by a prior probability distribution); Royall, *supra* note 34, at 173 (same).

the basis of observed evidence—and evidence alone—likelihood “supplies a natural order of preference among the possibilities under consideration.”⁴¹

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. The size of the likelihood ratio reflects the strength of this support. In fact, the size of the likelihood ratio arguably summarizes *all of the information* contained in the entire body of observed evidence.⁴²

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,⁴³ and so burdens of persuasion are used to categorize evidence that is *good enough* to legally establish still 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 today 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

⁴¹ RONALD A. FISHER, *STATISTICAL METHODS AND SCIENTIFIC INFERENCE* 68 (1956) (quoted in Edwards, *supra* note 30, at 27).

⁴² See, e.g., Edwards, *supra* note 30, 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).

⁴³ See McBaine, *supra* note 12, 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.”).

the fact-finder to “be persuaded by the evidence that the claim [or affirmative defense] is more probably true than not true.”⁴⁴ But it is also explained as requiring “any underlying material fact [to be] more likely [true] than not.”⁴⁵ As well as requiring that “the scales tip, however slightly, in favor of the party with [the] burden [of persuasion].”⁴⁶

In most criminal actions, due process demands a more stringent standard: proof *beyond a reasonable doubt*.⁴⁷ Explanations of this burden 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.”⁴⁸ 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.⁴⁹ Some jurisdictions refuse to attempt any definition at all.⁵⁰

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,

⁴⁴ Ninth Circuit, Manual of Model Civil Jury Instructions § 1.6 (2017) (brackets in original), http://www3.ce9.uscourts.gov/jury-instructions/sites/default/files/WPD/Civil_Instructions_2017_9.pdf; *see also* Seventh Circuit, Federal Civil Jury Instructions of the Seventh Circuit, *supra* note 10, at § 1.27 (“you must be persuaded that it is more probably true than not true”). *Cf.* Third Circuit, Model Civil Jury Instructions, *supra* note 10, at § 1.10 (“what [the plaintiff] claims is more likely so than not so”).

⁴⁵ *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”).

⁴⁶ *Ostrowski v. Atlantic Mut. Ins. Cos.*, 968 F.2d 171, 187 (2d Cir.1992).

⁴⁷ *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.”).

⁴⁸ Ninth Circuit, Manual of Model Criminal Jury Instructions § 3.5 (2010), http://www3.ce9.uscourts.gov/jury-instructions/sites/default/files/WPD/Criminal_Instructions_2017_9.pdf.

⁴⁹ *McBaine*, *supra* note 12, at 257 (summarizing these and other expressions).

⁵⁰ *See, e.g., United States v. Glass*, 846 F.2d 386, 387 (7th Cir. 1988) (“This case illustrates all too well that attempts to explain the term ‘reasonable doubt’ do not usually result in making it any clearer to the minds of the jury. And that is precisely why this circuit’s criminal jury instructions forbid them. ‘Reasonable doubt’ must speak for itself.”) (internal markup and citations omitted).

nor as well understood.⁵¹ 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.⁵²

Centuries of efforts to clarify the burdens of persuasion with linguistic refinements having failed,⁵³ the past 50 years have instead seen courts and scholars turn to probability concepts in their efforts to say what these standards mean. Edward Cheng recently summarized the probability thresholds that are often used to explain the most 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 0.95.⁵⁴

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) they would be more like 1.0 for preponderance and 9.0–19.0 for reasonable doubt, but the idea is the same.⁵⁵ Intermediate standards are less often targeted for quantification, and presumably fall somewhere between these extremes.

Many scholars harbor serious doubts about the Bayesian approach to fact-finding,⁵⁶ but with no alternative theory to take its place, the

⁵¹ 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”).

⁵² See McBaine, *supra* note 12, at 246 (arguing, in 1944, that “We should not ... have uncertainty, conflict and confusion in our legal system as to what are [the three common burdens of] persuasion and belief or how they should be adequately expressed in instructions for a jury.”).

⁵³ MCCORMICK ON EVIDENCE, *supra* note 51, at 662–63 (commenting on wasted effort arguing over linguistic metaphysics and “word-magic”).

⁵⁴ Cheng, *supra* note 15 at 1256.

⁵⁵ 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.

⁵⁶ See generally, Allen, *supra* note 17; Allen & Stein, *supra* note 29.

probability approach has long dominated academic research.⁵⁷ Courts are slowly adopting this approach too. While most continue to resist quantifying the reasonable doubt standard,⁵⁸ some now explicitly endorse a 50% probability threshold when explaining preponderance of the evidence to jurors.⁵⁹ And surveys of judges show rough consensus that preponderance of the evidence requires greater than a 50–60% probability of a fact's truth,⁶⁰ while reasonable doubt requires something higher, like a 75–90% probability of truth.⁶¹

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 novel 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.⁶² In 2013, Cheng proposed to model fact-finding in terms of

⁵⁷ 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).

⁵⁸ 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).

⁵⁹ 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.”).

⁶⁰ See, e.g., *U.S. v. Fatico*, 458 F.Supp. 388, 409–10 (1978) (citing and surveying evidence from two such surveys).

⁶¹ See, e.g., Posner, *supra* note 28, 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).

⁶² Kaplow, *supra* note 23, at 748 (distinguishing current law, which “takes behavior as given ... and asks, in light of that behavior, what is the likelihood of [harmful or

posterior probability ratios, but under *ad hoc* assumptions that happened to make these Bayesian probabilities behave like likelihoods in many respects.⁶³ 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.⁶⁴ Suffice it to say, 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.

3 A LIKELIHOOD THEORY OF FACT-FINDING

It is hard to think clearly about fact-finding without the foundation of a framework and lexicon for how facts, evidence, and causes of action interact in the justice system.⁶⁵ This is not to say that any of these terms are the least bit alien, only that they are rarely defined with precision in legal research. Even a tentative effort to standardize terms suggests the proposed likelihood theory of fact-finding.

3.1 *The General Fact-Finding Framework*

Reduced to core concepts, legal fact-finders are asked to compare uncertain facts to the elements of a cause of action in reaching a verdict. They do so by reference to evidence put forth by litigants, and subject to whatever burden of persuasion represents sufficient evidence to

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.

⁶³ Cheng, *supra* note 15, 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).

⁶⁴ 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...”).

⁶⁵ Oddly, research on fact-finding often skips this important and seemingly obvious step. *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).

prove the facts in a given context. None of this is new ground, but it is still instructive to take the uncommon step of considering each of the components of this framework in turn.

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. The universe and granularity of the possible facts is an empirical question of how a case is presented to and perceived by the fact-finder.⁶⁶

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 which the combination of facts makes out a cause of action.⁶⁷ 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

⁶⁶ See Michael S. Pardo & Ronald J. Allen, *Juridical Proof and the Best Explanation*, 27 LAW & PHIL. 223, 233–39 (2008) (providing a detailed and persuasive descriptive account of the types of factors that may help influence fact-finders' perception of the choice-space of material facts, including the *legal* elements of the cause of action, and the *focalizing* points of contrast between the parties' versions of events).

⁶⁷ 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.

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,⁶⁸ presence or absence of records of a regularly conducted activity,⁶⁹ etc. Some of it is evidence presented to build a broader narrative or story.⁷⁰ 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 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.⁷¹ 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,⁷² 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.

⁶⁸ *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).

⁶⁹ *E.g.*, FED. R. EVID. 803(6)–(7) (presence and absence of certain records).

⁷⁰ *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 65, at 662–65 (discussing the role of stories and “gaps” in evidence in a probability theory of fact-finding).

⁷¹ *E.g.*, FED. R. EVID. 801–807 (hearsay); FED. R. EVID. 602 (personal knowledge), FED. R. EVID. 403 (balancing probative value against risk of undue prejudice, etc).

⁷² *E.g.*, FED. R. EVID. 901.

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.⁷³ 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: something the fact-finder knows or can assess. This assumption is basic to a Bayesian probability approach to fact-finding as well,⁷⁴ and is actually more restrictive in that context than it is here: probability reasoning generally requires the fact-finder to be able to assess the absolute probability of seeing the evidence, whereas likelihood reasoning only requires the ability to assess the *comparative* probability of seeing the evidence under different factual hypotheses.

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.⁷⁵ To oversimplify this important

⁷³ See, e.g., Allen & Stein, *supra* note 29, 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 27, 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."); see also Uviller, *supra* note 22, at 846 ("As we assemble our evidence, we usually adopt the footprint theory of how to show who passed by. According to this theory, the past is preserved in the present by altered surfaces of matter and mind.").

⁷⁴ This assumption has undergirded the Bayesian analysis of fact-finding since the earliest works on this topic. E.g., Kaplan, *supra* note 27, 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.").

⁷⁵ See, e.g., *supra* notes 17–19; see also Allen & Stein, *supra* note 29, at 567–71 (discussing and citing literature related to the relative plausibility model).

body of work, fact-finders appear to assess the merits of competing factual narratives by comparing the epistemic credentials of different factual theories under the evidence. The type of factors considered in this process include the coherence, consilience, and causal specificity of stories about the facts, and other assessments of the comparative probability of observing the evidence under different factual hypotheses.⁷⁶ Put another way, the empirical research suggests that fact-finders interpret evidence in relation to factual narratives in a way that is very difficult, if not impossible, to distinguish from likelihood reasoning.

3.2 The Likelihood Approach to Fact-Finding

A 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 that 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 the current legal fact-finding process.

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. The Law of Likelihood provides that the observed value of a random variable (evidence) is evidential

⁷⁶ See Allen & Stein, *supra* note 29, 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.”).

support for one hypothesis (factual theory) over another (opposing 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 more demanding evidentiary requirements for the top hypothesis to be accepted—i.e., 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. This ratio thus compares the likelihoods of the single *most* plausible factual theory favoring each side of the case.

The 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.⁷⁷ 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.⁷⁸

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 this makes a great deal of sense. First, to be clear, this does *not* prevent either party from arguing in the alternative; it simply means that alternative arguments will be considered individually, and not be aggregated in the analysis. Second, the non-aggregation of alternative hypotheses is not artificial or imposed on the framework, but a substantive implication of the fact that prior probabilities do not factor into

⁷⁷ 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).

⁷⁸ Allen & Stein, *supra* note 29, 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.”).

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 which aggregation over alternative hypotheses is not generally possible or coherent.⁷⁹ Third, the ultimate choice between two fully specified factual theories is an intuitively desirable property of *any* theory of legal fact-finding. It means that fact-finding will always result in a single, specific finding of fact—something that an approach based on the aggregation of alternative factual hypotheses could never 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, 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 threshold value of the preponderance standard—and of every other standard, for that matter—requires a consideration of the underlying social and legal objectives behind the burden. Most discussion of the preponderance standard is an unhelpful mess of verbal gymnastics⁸⁰ and imprecise notions of evidentiary weight.⁸¹ But from this general confusion, two clear principles do emerge.

First, the preponderance standard places no special weight on the direction of any mistakes that the fact-finder may 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

⁷⁹ See *supra* notes 33–36 and accompanying text. Intuitively, without the basis of a prior probability distribution, it is impossible to say how much weight to assign each likelihood in an attempted aggregation. The lack of prior distribution does not mean that each possibility has equally prior probability, as would be needed, for example, to simply add up the various raw likelihood of alternative hypotheses.

⁸⁰ See *supra* note 53.

⁸¹ See, e.g., MCCORMICK ON EVIDENCE, *supra* note 51, at 660 n.7 (noting common judicial instruction against simply adding up the number of witnesses).

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.⁸²

Similar assertions are often made by scholars commenting on the basis for the preponderance standard.⁸³

Second, no particular strength of evidence is needed to meet this burden. Sometimes described as the “greater weight of the evidence,”⁸⁴ sometimes as facts just “more likely than not” in the plaintiff’s favor,⁸⁵ the inquiry is consistently into the *direction* that the evidence points, and not into how strongly it points in that direction.⁸⁶

These two principles translate into the likelihood theory of fact-finding 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.⁸⁷ 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 necessarily gives 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

⁸² *In re Winship*, 397 U.S. 358, 371–72 (1970) (Harlan, J., concurring).

⁸³ *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 51, at 669 (making a similar assertion).

⁸⁴ *E.g.*, *United States v. Matlock*, 415 US 164, 178 n.14 (1974).

⁸⁵ *E.g.*, *St. Mary's Honor Center v. Hicks*, 509 US 502, 527–28 (1993).

⁸⁶ *But cf.* McBaine, *supra* note 12, at 238–39 (noting that some courts that have tried to distinguish relative weight-of-evidence from absolute persuasion in this context).

⁸⁷ *Supra* note 46.

candidate theory favoring one of the two parties,⁸⁸ the winning story under the preponderance standard is, by definition, the *most* likely factual story on the evidence.⁸⁹ Put another way, this likelihood theory reduces the preponderance-of-the-evidence standard to this question: “What is the *most* plausible combination of facts on the observed evidence?” The party that the answer favors wins.

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.⁹⁰ While asymmetric error tolerance is implicit in these contexts,⁹¹ it is less emphasized than the heightened evidentiary requirement placed on the party bearing the burden.⁹²

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

⁸⁸ 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 .

⁸⁹ This assumes unique suprema. Otherwise multiple factual stories might be equally most likely, which complicates discussion of the likelihood theory, but little else.

⁹⁰ See, e.g., MCCORMICK ON EVIDENCE, *supra* note 51, 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”).

⁹¹ 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”).

⁹² E.g., MCCORMICK ON EVIDENCE, *supra* note 51, at 659–60 (commenting that unlike reasonable doubt, the preponderance and clear-and-convincing formulations direct attention to the evidence); McBaine, *supra* note 12, at 253 (commenting similarly).

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.⁹³

None of this says what the numerical value of k is—and it may well differ by type of case and jurisdiction.⁹⁴ 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 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,⁹⁵ 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 is often claimed that the reasonable-doubt standard differs in kind from the preponderance and clear-and-convincing standards; where the latter two focus on the *weight of the evidence*, the former is said to look to the mental state or *belief* of the fact-finder.⁹⁶ This seems more a point of form than substance. The only clear social and legal distinction drawn 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:

⁹³ Obviously, if the defendant bears this burden of persuasion on a particular issue, then the logic is reversed.

⁹⁴ See *supra* note 51.

⁹⁵ See, e.g., Royall, *supra* note 34, at ch. 1 (providing several examples).

⁹⁶ See, e.g., McBaine, *supra* note 12, 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 51, at 659–60 (making a similar observation).

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.⁹⁷

Similar statements have motivated the Court's decisions in other cases involving this standard,⁹⁸ and the overarching concern with false convictions is widely endorsed in the broader legal community as well.⁹⁹

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 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.¹⁰⁰ 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 comparing any two hypotheses, f' and f'' ,¹⁰¹ where f'' represents the true historic facts, the probability of seeing strong evidence that misleadingly favors f' over f'' is *always* bounded from above by the inverse of the strength of the evidence:¹⁰²

⁹⁷ In re Winship, 397 U.S. 358, 372 (1970) (Harlan, J., concurring).

⁹⁸ 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 erroneous [conviction]. In the administration of criminal justice, our society imposes almost the entire risk of error upon itself”).

⁹⁹ E.g. MCCORMICK ON EVIDENCE, *supra* note 51, 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.”).

¹⁰⁰ One might worry that extremely strong evidence could represent other concerns, such as the destruction of bad evidence, or appropriation of false testimony. This is a fair point, but in principle would be built into the weighing of evidence by the fact-finder, and so does not require any special treatment as a theoretical matter.

¹⁰¹ 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.

¹⁰² Royall, *supra* note 34, at 7 (proving and illustrating this universal bound).

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

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

This is not exactly a statement about controlling the global probability of false conviction as it cannot always be assumed that either f' or f'' is true in a given context,¹⁰³ but it does formalize the 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.¹⁰⁴ That is, this bound might accord the criminal defendant more advantage than intended. 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 necessarily 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

¹⁰³ 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).

¹⁰⁴ 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 34, 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 15, at 1277–78.

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; equating to at least a 90% probability of guilt. As already noted, the likelihood approach does not treat guilt or innocence as random in a given case. Instead, the *process* of using this likelihood ratio test, with this threshold value, would—on average, and in the long run—yield no more than 10% strong and misleading comparisons.

3.3 Intuition for the Likelihood Approach

This likelihood theory of fact-finding reduces every burden of persuasion to the same rule of likelihood reasoning. That rule can be further reduced to a four-step algorithm for deciding any case.¹⁰⁵

- 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 on the evidence.
- Fourth, decide for the plaintiff *if and only if* the relative likelihood of the set of facts favoring the plaintiff—the weight of evidence—is strongly enough in the plaintiff’s favor, meaning
 - *evidence weighs in favor of the plaintiff* (in theory, a likelihood ratio greater than 1) for preponderance of the evidence;
 - *evidence weighs strongly in favor of the plaintiff* (in theory, a likelihood-ratio threshold somewhere between 1–10) for clear-and-convincing evidence;

¹⁰⁵ For simplicity of presentation, this algorithm abstracts from things like affirmative defenses, for which the plaintiff might lose despite making out a cause of action. The approach can be easily extended to encompass these nuanced applications.

- *evidence weighs very strongly in favor of the plaintiff* (in theory, a likelihood-ratio threshold greater than 10) for the beyond-a-reasonable-doubt standard.

Put another way, the party with the burden of persuasion prevails when *the weight of the evidence* falls strongly enough in that party's favor. This is an objective inquiry about the relative consistency of the two competing factual theories with the observed evidence. In no event is the fact-finder required—or permitted—to base the verdict on what they personally *believe* the facts to be.

This theory of the fact-finding process differs fundamentally from the conventional Bayesian probability or belief-based account of fact-finding. Unlike probability, likelihood is a purely relative concept of evidential support.¹⁰⁶ The fact-finder is not tasked with forming propositional beliefs about the truth or accuracy of any given combination of facts in isolation in likelihood reasoning. Rather, every step in the process involves the simple *comparison* of two alternatives in terms of their consistency with the evidence. This mirrors arguments in the broader evidence literature that juridical fact-finding is a process of exclusively comparative reasoning—not reasoning from the absolute truth of any given factual hypothesis in isolation.¹⁰⁷

In fact, the proposed likelihood theory can be seen as formalizing much of the relative plausibility, narrative coherency, and story-based models of fact-finding, each of which describes some form of iterative

¹⁰⁶ See, e.g., Edwards, *supra* note 30, 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 34, 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.”).

¹⁰⁷ 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 15, 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.”).

comparison of alternatives in an effort to find the *relatively* most plausible story on the observed evidence.¹⁰⁸ But the likelihood theory also adds to these cognitive models of fact-finding. First, it explains how these models scale beyond the preponderance-of-the-evidence setting by providing a strength-of-evidence concept applicable to heightened burdens of persuasion. Second, it justifies what has been supposed a limitation of these models—their focus on the comparison of holistic factual stories rather than element-by-element evaluation against the burden of persuasion. Like these models, the proposed likelihood theory of fact-finding involves the comparison of holistic factual stories. The reason it does so is instructive: since the likelihood of any single item of evidence can depend on anything in the entire set of hypothesized facts, there is generally no way to separately evaluate isolated facts or elements in a likelihood approach to fact-finding.

A final corollary of this reliance on purely comparative reasoning is that the fact-finder must always consider a no-remedy factual theory in order to decide a case. This does not limit the defendant's right to hold the plaintiff to her proof; nor does it mean that the defendant must put forth a specific theory to prevail.¹⁰⁹ But even if the defendant does not put forth a specific no-remedy theory, the fact-finder must consider one, by independent inference if nothing else. This seems surprising at first, and inconsistent with common ideas about how fact-finding works in criminal settings, but it is squarely consistent with the even more fundamental idea that fact-finding should always result in a finding of facts. The final step in this likelihood theory is a choice between two fully specified factual theories; to decide for either party is to find the

¹⁰⁸ See generally *supra* notes 17–19 (citing works in this literature); see also 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 17, 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 28, at 1513 (“[The] benchmark for the plaintiff’s case is not the null hypothesis but the defendant’s case.”).

¹⁰⁹ Cf. Cheng, *supra* note 15, 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.’”).

factual theory favoring that party to be *legally true* under the applicable burden of persuasion.

4 COMPARISONS AND IMPLICATIONS

Perhaps the most fundamental thing one might demand of a theory of fact-finding is that it should provide sound suggestions for idealized fact-finding—giving the theory the normative content to say how trial practice and the rules of evidence *should* handle various situations. Decades of work on probability and belief-based theories have failed to produce a satisfying description of the ideal fact-finding process.¹¹⁰ But by switching focus away from beliefs about the probabilistic truth of facts, and toward the relative consistency of factual hypotheses with the evidence, the proposed likelihood theory of fact-finding fares far better. This section highlights some preliminary implications of this change in focus, addressing two questions in particular:

- Which approach—probability or likelihood—is the more plausible description of the fact-finding process?
- What normative implications does this new likelihood theory have for trial practice and the rules of evidence?

The answers to these questions touch everything from the mechanical process of the ideal fact-finder, to the language of jury instructions, to the justification of evidentiary resistance to character and propensity reasoning and naked statistical evidence, to the conceptual framework for explaining how prejudice and personal bias affect fact-finding and *why* just fact-finding cannot their influence.

4.1 *Implications for the Ideal Fact-Finding Process*

To show how the proposed likelihood theory differs from probability and belief-based models of the ideal fact-finding process, it suffices to survey a few of the many difficulties inherent in these probability and belief-based models. Indeed, even the basic mechanics of a Bayesian theory of fact-finding are surprisingly illusive. The problem is that all

¹¹⁰ Compare Allen, *supra* note 107, at 402–03 (noting problems with probability theories of fact-finding in the 1980s) with Allen & Stein, *supra* note 29, at 560–65 (noting the same problems and paradoxes in modern probability theories of fact-finding).

probability statements describe something about the *total* uncertainty in a system; the probabilities of all possible alternatives sum to one. To see the headaches this creates for a probability or belief-based description of idealized fact-finding, consider even the lowly preponderance-of-the-evidence standard.

The traditional probability-based articulation of the preponderance standard states that the plaintiff must prove “the existence of [any material] fact [to be] more probable than its nonexistence.”¹¹¹ Familiar jury instructions use almost exactly these terms.¹¹² Put more formally, the plaintiff is required to prove that the probability of a combination of facts in the cause-of-action set is more than 50% or—equivalently—that the probability that this combination of facts is true is greater than the negation of that claim.¹¹³ The suggestion is that the idealized fact-finder would assess the posterior probability of the plaintiff’s theory on the evidence, then assess the probability of *all other possible combination of facts* on the evidence, and then compare the ratio of these probabilities to 1 to see if the plaintiff wins.

This leads to shockingly absurd results. Suppose two cars collide on a 25-mph road, and that the plaintiff’s theory includes a claim that the defendant was driving 60 mph at the time of the collision. The negation of this claim includes *every other speed* that the defendant could have been driving—including still dangerous speeds (e.g. 59.5 mph) as well as even more dangerous speeds (e.g. 70 mph). A strict application of the above test would count the probabilities of these strongly pro-plaintiff alternative facts *against* the plaintiff’s right to recovery.

In fairness, this is a bit of a straw man: no advocate of probability reasoning would endorse the literal application of the traditional test, or the standard jury instruction that accompanies it. Instead, it might be

¹¹¹ Flemming James, *Burdens of Proof*, 47 VA. L. REV. 51, 54 (1961); see also McBaine, *supra* note 12, at 260–61 (suggesting a similar rule).

¹¹² See, e.g., *supra* note 44 and accompanying text.

¹¹³ 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 15, 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).”).

argued that the plaintiff is really alleging a composite *set* of alternative facts to support recovery. Perhaps the composite fact is that the defendant was driving “over the speed limit” in the prior example. But this only shifts the problem to a different place. Now the possibility that the defendant was not speeding, but was instead drunk or distracted with the radio counts against the plaintiff’s right of recovery in this idealized model of the fact-finding process.

The only way that aggregation of alternative factual theories fully escapes these bizarre results is for the idealized fact-finder to consider *every possible* combination of facts in the *entire* cause-of-action set in assessing the plaintiff’s right to recover. I am unaware of any common jury instructions to this effect, which is gratifying given the magnitude of the task it contemplates. The idealized fact-finder would need to (1) identify every combination of facts in the entire cause-of-action set, (2) assess the absolute probability of every combination of facts in this possibly infinite set, and (3) sum up all of these probabilities in arriving at a single aggregate probability or belief measure that would then be compared to the burden of persuasion.

As a description of idealized trial fact-finding, this seems wanting. First, the scope of the task feels out of place in the typical trial context. When—if ever—would trial map out the absolute probability of every possible combination of facts in entire cause-of-action set?¹¹⁴ Second, the approach leads to a peculiar concept of fact-finding. Suppose that each of six alternative factual theories is found to have only a 10% probability of being true. Together, the aggregate probability of these six alternative theories satisfies the preponderance standard, and thus entitles the plaintiff to relief. But why? What happened to justify relief? That is, what *facts* has the fact-finder *found* in this situation?

Seeking to avoid these problems, some scholars have abandoned aggregation strategies in favor of descriptions of idealized fact-finding based on the probability ratios of individual pairs of factual theories.¹¹⁵ Cheng provides a recent example. Though continuing to interpret fact-finding in terms of Bayesian probabilities, Cheng imposes the *ad hoc*

¹¹⁴ Cf. James, *supra* note 111, 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 . . .”).

¹¹⁵ Allen, *supra* note 107, at 425–28 (providing what appears to be the first formal suggestion of this approach).

assumption that fact-finders cannot aggregate the probabilities of alternative factual theories together, leaving only the comparison of individual factual theories as a basis for inference.¹¹⁶ If the plaintiff and defendant each advance one factual theory, Cheng's proposed model of idealized fact-finding defines the plaintiff's right to recover by the degree to which the posterior probability of the plaintiff's theory exceeds the posterior probability of the defendant's theory. That is, the idealized fact-finder would compare the posterior probabilities of *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,¹¹⁷ 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 claim the light was red and green, respectively. If the fact-finder concludes that the posterior probability of a red light on the evidence is $P(R|e) = 0.4$ and that the posterior probability of a green light on the evidence is $P(G|e) = 0.3$, then on Cheng's theory the plaintiff wins and the defendant is liable. But since probabilities must always sum to one, the above conclusions necessarily imply the probability of a yellow light is $P(Y|e) = 0.3$. That is, Cheng's theory would be assigning liability under circumstances that actually prove the light was probably *not* red at all: $P(G \text{ or } Y|e) = 0.6$.

All of these problems arise from reliance on absolute probability or belief as the measure of legal fact-finding. And by merely changing the concept of uncertainty from probability to likelihood—the theoretical analog of relative plausibility or relative consistency of different factual hypotheses with the evidence—all of these problems disappear.

- 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

¹¹⁶ Cheng, *supra* note 15, 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.”).

¹¹⁷ Cf. Allen & Stein, *supra* note 29, 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.”).

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 aggregate probability models, the likelihood theory involves a series of purely relative comparisons; and, in theory, it always concludes with a single, fully specified, finding of facts.
- Compared to probability ratio models, likelihood analysis requires no modification of the intrinsic properties of likelihoods; and the yellow-light paradox does not apply as likelihoods do not sum to one (that the observed evidence is more consistent with a red light than a green light does not by itself imply anything about the consistency of evidence with a yellow-light hypothesis).¹¹⁸

Instead, likelihood analysis supports the simple model of idealized fact-finding summarized in Section 3.3, and described in greater detail in Section 3.2. As already noted, this process—which involves only the iterative comparison of alternative factual stories on the available evidence—closely mirrors not only the typical presentation of evidence in adversarial litigation, but also persuasive empirical accounts of the deliberative process of many fact-finders.¹¹⁹

4.2 Implications for the Conjunction Paradox

Another view of how a likelihood theory of fact-finding differs from a probability or belief-based theory is afforded by how each approach conceptualizes *facts* in the fact-finding process. As already noted, the

¹¹⁸ One might wonder why the fact-finder doesn't view the possible fact as binary: that is, was the light *red* or was it *not-red*? Nothing prevents this possibility; whether the fact-finder approaches the set of possible facts as *red*, *yellow*, *green* or *red*, *not-red* is an empirical question of how the exercise is framed. See *supra* note 66 (citing work on how the set of potential facts may be framed in practice). The likelihood approach to fact-finding accommodates either version—a nod to practicality. Where the perfect platonic fact-finder might identify the defendant's exact speed, the exact color of the light, the exact demands of prudence under the circumstances, flawed human fact-finders will inevitably discretize theoretically continuous variables and group together alternatives that are not so conceptually separable. This is simply a matter of how the fact-finding problem is framed: a point on which likelihood theory is itself silent. See *infra* notes 134, 137, 140, 174–175 and accompanying text (discussing other instances of framing on which the theory of likelihood analysis is silent).

¹¹⁹ See *supra* notes 17–19, 107–108 and accompanying text.

likelihood model typically assumes that facts are immutable constants: examples include the historic actions of the parties, or a person's true mental state at the time of a transaction. By contrast, probability models of fact-finding treat the facts of a case as random variables, implying that the facts must obey the laws of probability. This, in turn, leads to predictably strange results.

A modest example is a complication arising from the property that continuous random variables have infinitesimal probability of taking on any given value. To illustrate, suppose a factual theory includes the claim that the defendant was driving 60 mph on a particular highway. The probability that the defendant was really going exactly 60 mph is technically zero. Even if this is a close estimate, the truth might be 61 mph, 59.5 mph, or 60.001 mph—with enough evidence, any single guess can be almost surely disproved. This technical oddity muddies the model: it means that the probability of any combination of facts including a continuous variable (like driving speed) can never exceed 50%, or even 0% for that matter. The problem is mitigated by aggregating probabilities of alternative factual theories, but that has its own problems, as discussed in the previous subsection.

Another—and far more troubling—complication arising from the treatment of facts as random variables is what the literature has come to call the *conjunction paradox*. In brief, the conjunction paradox is the unsettling discovery that when material facts are treated as random variables, the threshold probability needed to prove any given fact (or element of a cause of action) may diverge from the threshold probability needed to meet the burden of persuasion on the overall claim for relief. A thought experiment 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, which seems like it would satisfy the usual jury instruction on the preponderance standard.¹²⁰ But the probability of *both* facts being true may actually be much smaller than either individual probability. For example, if f_1 and f_2 are independent

¹²⁰ See *supra* note 44 and accompanying text.

variables (such that the truth or falsity of one fact suggest nothing about the truth or falsity of the other), then the probability of both facts being true is $P(f_1 \text{ and } f_2 | e) = 0.7 \times 0.7 = 0.49$, 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.¹²¹ Second, it implies that the practical difficulty of satisfying a burden of persuasion should strongly depend on the dubiously meaningful number of elements or material facts in a claim.¹²² To illustrate that issue, *theft* has more elements than *murder* in most jurisdictions, yet few see a viable argument for subjecting the elements of theft to more a searching evidentiary standard.¹²³ 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 not just on that issue, but on all *other* issues as well.¹²⁴

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.¹²⁵ Or that alternative factual

¹²¹ Cf. Posner, *supra* note 28, at 1513–14 (oscillating on how a fact-finder would decide the case when individual and joint probabilities lead to different outcomes).

¹²² E.g., Sanchirico, *supra* note 26, 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 107, 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.”).

¹²³ Leiter & Allen, *supra* note 14, at 1504–05.

¹²⁴ See Cheng, *supra* note 15, 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.”).

¹²⁵ 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 29, at 595–96.

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).¹²⁶ But both of these arguments amount to *ad hoc* modification of the rules of probability in a theory meant to derive from the rules of probability—perhaps a hint that probability concepts are an awkward way to understand uncertainty in legal fact-finding.

Likelihood reasoning avoids all random-facts problems, including the conjunction paradox. It does so by not treating the facts as random variables in the first place. Instead, likelihood analysis generally treats the facts of a case as fixed, and the *evidence* as random. It then compares the probability of seeing the observed evidence under different factual hypotheses. The source of the conjunction paradox—the product rule of probabilities—still applies in likelihood analysis, but it applies to the *evidence*, not the facts. And since the same evidence appears on the top and the bottom of the likelihood ratio, the effects of the product rule essentially cancels out.

Intuitively, likelihood reasoning treats the facts of the case as the 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 does not have any general effect on the ease or difficulty of meeting the burden of persuasion. Two implications are apparent from this result: (1) it is generally inappropriate to describe burdens of persuasion as applying to individual elements of a cause of action, and (2) when approached in likelihood terms, fact-finding exhibits no conjunction paradox.

4.3 *Prior Probability, Character Reasoning, and Naked Statistics*

A third useful comparison between probability and likelihood theories of fact-finding, and their corresponding implications for fact-finding, is afforded by the different way these theories suggest that fact-finders should use prior probabilities in their deliberations. To keep

¹²⁶ Cheng, *supra* note 15, 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.

things concrete, this discussion anchors around two infamous puzzles in the evidence literature: the *Gatecrasher* and the *Blue Bus* paradoxes.

The Gatecrasher paradox is a toy fact-pattern that leads to a strange result under traditional probability reasoning.¹²⁷ 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 that a randomly chosen attendee would be one of the gate-jumpers is 50.1%, making any randomly chosen attendee liable for the price of admission under the preponderance standard of a Bayesian probability model of fact-finding. Everyone agrees that Bayesian logic compels this result, but no one seriously thinks it is the right outcome.¹²⁸

The Blue Bus paradox is a slightly different situation, loosely based on the facts of an actual case.¹²⁹ 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 up by either side. Many commentators conclude that this means the Blue Bus Company has an 80% probability of responsibility.¹³⁰ But, again, few are comfortable assigning liability on this record alone.¹³¹

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

¹²⁷ See L. JONATHAN COHEN, *THE PROBABLE AND THE PROVABLE* 75 (1977) (giving the first description and illustration of this paradox).

¹²⁸ E.g. Allen & Stein, *supra* note 29, 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.”).

¹²⁹ *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 20, at 1341–42, n.7.

¹³⁰ The standard *Blue Bus* hypothetical actually provides insufficient information to compute an unambiguous probability of responsibility. Cf. *infra* note 137.

¹³¹ E.g. Posner, *supra* note 28, 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).

drawing a negative inference from the absence of better evidence in both puzzles.¹³² 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 than technically needed to win in these scenarios.¹³³ Perhaps, but this is again dodging the core difficulty of each puzzle. Instead, the clearest way to reconcile the result in each of these puzzles with our intuition is to say that the *paradox* in each puzzle comes from the effort to base fact-finding conclusions on prior probabilities.

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:

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

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 equals precisely *one* in this puzzle: the only “evidence” is the contents of the ticket box, and the contents of this box would be identical 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 of the gate-jumpers. 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 a Bayesian probability theory of fact-finding.

¹³² *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 28, 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.”).

¹³³ *E.g.* Posner, *supra* note 28, at 1509 (arguing that judicial economy may require the plaintiff to expend more effort in these cases); *see also* Wells, *supra* note 57, at 740 (noting and citing various similar policy arguments).

This liability determination is entirely driven by prior probabilities, and discomfort with the outcome belies an intuitive resistance to assigning guilt or liability on a prior-probability basis. The principles of likelihood reasoning comport with that hesitancy. In fact, likelihood analysis of this puzzle does not consider prior probabilities at all. With a likelihood ratio of exactly one, there is not enough evidence (literally “no evidence”) to meet the preponderance standard of the likelihood theory of fact-finding. In contrast to the paradoxical result reached by Bayesian probability reasoning, a randomly chosen attendee would *not* be held liable under a likelihood approach to this puzzle.

This difference in outcomes reflects a fundamental disparity in the way each theory treats information about average behavior and background frequencies. As explained above, Bayesian probability analysis expects the fact-finder to base decisions—at least in part—on the prior probabilities of material facts. Here, the ticket box that describes the *average* behavior of all the rodeo attendees is used to reason about the culpability of a single, randomly chosen attendee. Likelihood analysis does not involve prior probabilities in any stage of the fact-finding process. Here, the average number of paying customers in the stands is simply irrelevant to a likelihood analysis of the puzzle. To see just how far likelihood reasoning goes in ignoring prior probabilities, consider the following two variations on the usual 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 hardly 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* suffices to carry the preponderance standard under likelihood reasoning. It does not matter that only one attendee in the crowd jumped the gate. The legal inquiry is about *this* attendant, and the only available evidence points toward liability.¹³⁴

¹³⁴ Some might puzzle why the inquiry is between crowd members and not between a crowd member and someone in the larger public: with even a single ticket missing, the likelihood that a randomly chosen attendee is a gate jumper will always exceed the likelihood that someone outside the stadium is a gate jumper. But this reflects no weakness of likelihood analysis. It simply reflects the fundamental problem that the framing of fact-finding matters. By emphasizing the importance of framing effects, the likelihood theory may help to clarify what the correct frame of reference actually is in the fact-finding process. Here, for example, the relevant legal question seems to

Second, suppose there is no testimony to be had, but a look in the ticket box reveals one lonely ticket—the implication being that fully 999 out of the 1,000 attendees have jumped the gate, making the prior probability 99.9% that a randomly selected attendee would be a gate-jumper. Just like the standard Gatecrasher puzzle, this version of the puzzle presents no *evidence* in the likelihood sense of the term. One attendee has paid the price of admission, and nothing has been shown to indicate that this particular attendee is more likely a jumper than the (one) payer. Thus, even in this extreme setting, likelihood reasoning would not find a randomly chosen attendee liable for the price of admission.¹³⁵ Of course, social policy might suggest holding *all* members of the group jointly liable in such an extreme situation.¹³⁶ But that only clarifies the critical point: available information proves *the group* liable, not any particular *individual* therein.

Similar logic applies in the Blue Bus case. The clearest way to frame this puzzle is as a prior probability ratio of 80%/20% in favor of the Blue Bus Company's responsibility, again with a likelihood ratio of exactly one.¹³⁷ As in the Gatecrasher puzzle, lack of likelihood-relevant evidence prevents a finding of liability under likelihood analysis. But consider a related puzzle proposed by Gary Wells: the same circumstances as before, but instead of the Company owning 80% of the blue

be *which* of the attendees jumped the gate (the between attendee comparison), not *whether* any of them did (the between attendee-and-non-attendee comparison).

¹³⁵ This is a knife-edge result, since a ticket box without a single ticket in it would conclusively prove *every* attendee liable. This discontinuity may initially seem odd, but as explained in the remainder of the paragraph, it makes sense when approached from the perspective of an inquiry about an *individual* defendant. Similar discontinuities already exist in the law of evidence. Compare FED. R. EVID. 404(b)(1) (stating the usual prohibition on the use of evidence of prior bad acts to infer repetition of the bad act on a particular occasion) with MCCORMICK ON EVIDENCE, *supra* note 51, at 1035–37 (summarizing the usually permitted use of evidence of highly distinctive prior acts—*modus operandi*—to infer repetition of the act on a particular occasion).

¹³⁶ 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).

¹³⁷ 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.

buses, a weigh-station logbook records a Blue Bus Company bus passing down the road just before the collision—although, on cross exam, the logbook is also shown to be accurate only 80% of the time.¹³⁸ This revised puzzle contains likelihood-relevant evidence. The logbook has an 80% probability of correctly identifying the Blue Bus Company if a Company bus was responsible for the incident, and only a 20% probability of falsely identifying the Blue Bus Company if the offending bus did not belong to the Company ($LR = 0.8/0.2 = 4$). A likelihood approach to fact-finding would thus assign liability in the logbook version of the puzzle ($LR = 4$), but not the canonical version ($LR = 1$).¹³⁹

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 the same 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.

These examples illustrate likelihood reasoning's sensitivity to the type of information being considered: likelihood-relevant evidence is used; prior probability is not. Admittedly, the distinction between these concepts is not always great, and whether something characterizes evidence or prior probability can depend on how fact-finding questions are framed.¹⁴⁰ But while resting a sharp distinction on potentially so

¹³⁸ Wells, *supra* note 57, at 741.

¹³⁹ Some may puzzle about this difference: how is an ownership proportion of 80% different from the testimony of a witness who is correct only 80% of the time? One way to consider the difference is to ask how the available information would differ if the Blue Bus company were *not* responsible. It would obviously own 80% of the blue buses either way, but the eye witness's testimony has only a 20% chance of falsely identifying the Blue Bus company if that company was innocent. Information that depends on the underlying facts is likelihood-relevant evidence.

¹⁴⁰ 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

subtle a difference may seem like a defect of the likelihood theory of fact-finding, it may actually be the theory's greatest strength.

First, whether ideal or not, the sharpness of this distinction reflects empirical realities. 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 analysis to the same 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-versions 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 of the puzzle, and rarely in the canonical version.¹⁴¹ This is inconsistent with fact-finding based on Bayesian probability reasoning: subjects assess the same posterior probabilities in both versions of the puzzle, but reach different results at the fact-finding stage. It is entirely consistent with likelihood reasoning—and with subjects demanding individualized, likelihood-relevant evidence to support a finding of liability.¹⁴²

Second, the distinction between individualized, likelihood-relevant evidence and prior probability reflects existing concepts in the law of evidence. One clear example is the usual ban on character reasoning: evidence of a person's character (or general propensity to act some way) is generally inadmissible to prove that the person did act in conformity with that character (or propensity) on a particular occasion.¹⁴³

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).

¹⁴¹ *Id.* at 742 (figure 1); *see also id.* at 744 (figure 3).

¹⁴² *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.”).

¹⁴³ *E.g.* FED. R. EVID. 404; *cf.* Paul F. Rothstein, *Intellectual Coherence in an Evidence Code*, 28 LOYOLA L. REV. 1259 (1995) (discussing and critiquing theories of the justification for this doctrine, and its exceptions); Peter Tillers, *The Death of a*

This is not an obscure or technical evidentiary exclusion. It is a fundamental rule of evidence that codifies the prior-probability distinction of likelihood analysis by prohibiting the fact-finder from using at least character-based prior-probabilities in deciding how a person actually acted on a particular occasion.¹⁴⁴

Another example is judicial hostility to *naked statistical evidence*. When faced with purely statistical evidence, courts and fact-finders often refuse to rely upon statistics that are not in some way individualized to the specific parties or transaction at issue in a case.¹⁴⁵ An old state-court opinion colorfully captures the 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

Youth and of a Drunkard: A Remarkable Story of Habit and Character in New Jersey, in EVIDENCE STORIES 29, 30 (Richard Lempert ed. 2006) (“Guilt or innocence is supposed to depend on what a person does, and not on what a person is.”).

¹⁴⁴ One might reason that rules of evidence allowing the use of evidence of habitual acts represent a counter-argument. See, e.g., FED. R. EVID. 406 (“Evidence of a person’s habit ... may be admitted to prove that on a particular occasion the person ... acted in accordance with the habit...”). But this, too, is consistent with likelihood analysis. The confusion is another subtle matter of framing. The prohibited act in character reasoning is to infer from evidence of a person’s propensity to act some way (prior probability), something about the conscious and willful (non-random) acts of that person. By contrast, the type of habitual-act evidence that is most clearly admissible is that describing the nearly automatic act of a person. See, e.g., FED. R. EVID. 406 Advisory Committee’s Note to the 1972 Proposed Rules ¶ 2 (“The doing of the habitual acts may become semi-automatic...”). These semi-automatic acts can be seen as random processes: facing certain specific conditions, the person exhibits a pattern response with a low probability of deviation. In the specific case where the fact at issue is itself an explicitly *random process*, likelihood analysis has no objection to prior probabilities. This is not a defect of the likelihood model, but a recognition of the different type of uncertainty inherent in random processes. The topic of truly random facts is discussed further in Section 5, below.

¹⁴⁵ 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 113, 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.”).

throw such was the actual result. . . . The slightest real evidence that sixes did in fact fall uppermost would outweigh all the probability otherwise.¹⁴⁶

This demand for individualized “evidence” over mere “probability” is bewildering from a Bayesian perspective, since *both* individualized evidence (likelihood-relevant evidence) and statistical averages (prior probabilities) factor into the computation of posterior probabilities.¹⁴⁷ But judicial insistence on individualized evidence is easily explained by the likelihood theory of fact-finding. The distinction between “evidence” and “probability” seems to be neither more nor less than the distinction between likelihood-relevant evidence and prior probability. In fact, efforts to define naked statistical evidence actually provide a useful rule-of-thumb for identifying information that characterizes prior probabilities: the information 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.”¹⁴⁸

Third, to the extent that the sharp distinction between likelihood-relevant evidence and prior probability reflects common fact-finding practice, it is important that a theory of fact-finding encompass this distinction. While it may often be clear 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 by far that the legal community recognize and struggle to enforce the difference between prior probability and likelihood-relevant evidence, than for this distinction to be swept under the rug as it would be in a Bayesian understanding of the fact-finding process.

¹⁴⁶ Day v. Boston & M.R.R., 96 Me. 207, 52 A. 771, 774 (1902).

¹⁴⁷ See 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”); see also Peter Tillers, *If Wishes Were Horses: Discursive Comments on Attempts to Prevent Individuals from Being Unfairly Burdened by Their Reference Classes*, 4 *LAW, PROBABILITY, & RISK* 33, 36 (2005) (presenting and discussing an argument “that the . . . distinction between specific and non-specific [statistical] evidence is almost unintelligible”).

¹⁴⁸ Wells, *supra* note 57, at 739 (providing this as a typical definition for “Naked statistical evidence,” but noting that the term is ill defined in the legal literature”).

4.4 *Personal Beliefs, Prejudice, and Biased Deliberations*

A final distinction to highlight between a probability or belief-based approach to fact-finding and a likelihood approach is the different tolerance each theory has for subjective beliefs and fact-finder bias. Because a probability theory of fact-finding relies on the fact-finder to supply his or her own prior probabilities for all material facts, it necessarily and explicitly builds subjective prior beliefs, prejudice, and bias into its account of the idealized fact-finding process. Likelihood reasoning contemplates no such thing. Idealized fact-finding in the proposed likelihood theory is as free of fact-finder bias as it can be. The likelihood theory provides a framework for understanding the harm of using prior beliefs in fact-finding, and also a normative argument—as if one were needed—for the elimination of all forms of subjective prior belief, prejudice, and bias in fact-finding.

To see how deeply and distressingly prior beliefs shape fact-finding in a probability or belief-based approach, consider how a Bayesian model of the fact-finding process translates into likelihood terms.¹⁴⁹ Suppose that only two factual theories are at issue. The first, $f' \in C$, is the plaintiff's theory of the case; the second, $f'' \in C^c$, is the defendant's theory. In a Bayesian approach, the idealized fact-finder would find for the plaintiff if the posterior probability of the plaintiff's theory sufficiently exceeded that of the defendant's theory:

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

The above formula is simply Bayes' Theorem with the likelihood-ratio (middle term) expressed in likelihoods instead of probabilities. The threshold, m , represents the burden of persuasion.¹⁵⁰

By rearranging terms, the above Bayesian probability test can be expressed in likelihood-ratio terms:¹⁵¹

¹⁴⁹ This presentation of posterior probability reasoning in terms of likelihood ratios is similar to that of Kaplow, *supra* note 64, at 34–35.

¹⁵⁰ See *supra* notes 54–55 and accompanying text.

¹⁵¹ 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.

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

That is, a Bayesian or belief-based theory of fact-finding is equivalent to a specific form of likelihood reasoning in which the relative weight of evidence needed to justify conviction or liability depends on the fact-finder's prior beliefs. Specifically, the probability or belief-based approach is similar to the proposed likelihood theory of fact-finding, but with a burden of persuasion defined as a fixed constant (the formal burden of persuasion) that is then scaled up (raising the burden) or scaled down (lowering the burden) to reflect the fact-finder's personal beliefs about the facts at the start of a case.

The idea that a burden of persuasion would depend on the personal views of the fact-finder is troubling for at least two reasons. First, in contrast to some fields of study—where the influence of prior beliefs might be washed away by the collection of enough data—there is no reason to suspect that prior beliefs will be so diluted by the evidence in a typical fact-finding exercise.¹⁵² Second, while there is nothing inherently debatable about how a fact-finder feels at the start of a case, the normative desirability of building these personal feelings into the fact-finding process is another matter altogether.¹⁵³ It seems fair to insist that the underlying epistemology of legal fact-finding should be a public concept, not a private one.¹⁵⁴ If this is too abstract, a concrete example easily illustrates the point.

The introduction of this paper briefly touched upon the recent Supreme Court case of *Peña-Rodriguez v. Colorado*. That case concerned

¹⁵² See Leiter & Allen, *supra* note 14, 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.”).

¹⁵³ See Pawitan, *supra* note 33, 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.”).

¹⁵⁴ Cf. Taper & Lele, *supra* note 103, 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.”).

a juror who explained during deliberations that the Latin-American defendant, on trial for sexual harassment, could be presumed guilty 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 whatever they wanted with women.”¹⁵⁵ The juror went on to share his view 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.”¹⁵⁶ As if to dispel any doubt that these statements were meant to describe the *prior probability* of the defendant’s guilt, the juror concluded that “in his experience, ‘nine times out of ten Mexican men were guilty of being aggressive toward women and young girls.’”¹⁵⁷ Suppose, for sake of argument, that these (sad and ignorant) beliefs were sincerely held.

Any fact-finding process that is based on the fact-finder’s personal beliefs about the facts necessarily starts from the fact-finder’s prior beliefs about the facts—including prior beliefs that are, in turn, based on racial bias, gender stereotypes, assumptions about religious groups, and the like. A Bayesian theory of fact-finding not only condones, but actually *requires* reliance on these personal prior beliefs.¹⁵⁸ And in so doing, it prescribes the differential treatment of defendants. Applied to the situation in *Peña-Rodriguez*, for example, a probability or belief-based approach to fact-finding would instruct the biased juror to consider only the evidence presented at trial, but to demand a smaller quantum of evidence in order to convict a Mexican defendant than to convict an otherwise identical white (presumably) 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 much confidence in the use of

¹⁵⁵ *Peña-Rodriguez v. Colorado*, 137 S. Ct. 855, 862 (2017).

¹⁵⁶ *Id.*

¹⁵⁷ *Id.*

¹⁵⁸ See Cheng, *supra* note 15, at 1267 (“In legal venues, one might fear that [prior probability ratios] embody prejudices against certain types of parties.”); cf. Posner, *supra* note 28, 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.”).

prior beliefs in fact-finding.¹⁵⁹ The problem is that there is no way to avoid these implications and still retain a *probability* interpretation of the fact-finding process.

Cheng, for example, has attempted to argue that prior probability ratios always default to *one* in a fact-finding setting: “the legal system imposes a constraint ... [that normatively fixes] the prior odds ratio at 1 to start the plaintiff and the defendant in equipoise.”¹⁶⁰ But does this make sense? Cheng’s justifications for the claim are that it equates to unbiased fact-finding,¹⁶¹ is (somehow) implied by the plaintiff’s satisfaction of the burden of production,¹⁶² and is what fact-finders ought to do, even if they don’t do it.¹⁶³ These arguments are uninspiring, but even if they were, the use of 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 prior 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 subvert the legal process to act on his bias; it is simply

¹⁵⁹ *But cf.* Posner, *supra* note 28, 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.”).

¹⁶⁰ Cheng, *supra* note 15, at 1273.

¹⁶¹ *Id.* at 1267 n.24 (citing Posner, *supra* note 28, at 1514).

¹⁶² Cheng, *supra* note 15, 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.

¹⁶³ *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).

a truism that this person 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, at the start of a case, “I do not have any idea whether the defendant is a murderer” is obviously *not* the same as saying “I believe there is a 50% probability that the defendant is a murderer.”¹⁶⁴ To further illustrate that point, suppose a defendant is on trial for two independent charges of murder, and that you know nothing about the defendant’s guilt on either charge. Reflecting this initial-state ignorance via a prior probability guilt of 50%, means—surprisingly—that you must believe there is a 75% probability that the defendant committed at least one of the two murders. The attempt to represent initial-state ignorance with uniform prior probabilities about one aspect of the case has translated into gratuitous information about another aspect.¹⁶⁵

The likelihood theory of fact-finding escapes all of these problems. Likelihood reasoning has no use for personal prior beliefs, prejudice, or bias of any form. It is reasoning from evidence alone. Put another way, likelihood analysis is the evidence-theory answer to the question Justice Sotomayor posed during oral arguments for *Peña-Rodriguez*, quoted in the first sentence of this paper: “Don’t we want deliberations on evidence and not deliberations on someone’s stereotypes and feelings about the race of a defendant?”¹⁶⁶ Of course we do. Likelihood reasoning encapsulates precisely this approach to fact-finding.

The proposed likelihood theory of fact-finding eliminates the unsettling notion that different defendants should face different effective burdens of persuasion. In the likelihood approach to fact-finding, the idealized fact-finder is always comparing the same evidence to the same threshold, regardless of the fact-finder’s personal feelings and convictions. Theories of fact-finding under uniform prior probabilities

¹⁶⁴ See Royall, *supra* note 34, 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.”).

¹⁶⁵ See Edwards, *supra* note 30, at 58.

¹⁶⁶ Transcript of Oral Argument at 40:5–7, *Pena-Rodriguez v. Colorado*, 137 S. Ct. 855 (2017) (No. 15-606).

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 has been looking for.

5 CONCLUSION

In 1944, J.P. McBaine wrote simply, but powerfully, that then-existing confusion about fact-finding was undermining the legal system:

[No legal system] can be praised for practicability if there exists vagueness, uncertainty or confusion as to the scope or extent of the burden [of persuasion], 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.¹⁶⁷

If this is the test, then we must concede that, nearly 75 years later, our legal system *still* cannot be praised for practicability. McBaine, like the many writers to follow, thought that probability concepts and a focus on the personal beliefs of fact-finders were the keys to untangling legal fact-finding. But as this article argues, probability and belief concepts are generally ill-suited to the task. I argue that to understand legal fact-finding and burdens of persuasion, our focus must be on the relative plausibility of competing narratives,¹⁶⁸ the relative weight of the evidence,¹⁶⁹ and the implications of likelihood reasoning.¹⁷⁰

The proposed likelihood theory of fact-finding makes ground on several fronts. It unifies understandings of the fact-finding process by providing a framework in which modern cognitive models of the fact-finding process extend to all burdens of persuasion. It also unifies the procedural theory of fact-finding with the law of evidence, coinciding with evidence-law principles such as the ban on character reasoning¹⁷¹

¹⁶⁷ McBaine, *supra* note 12, at 242.

¹⁶⁸ See *supra* note 17 and accompanying text (discussing empirical research on the cognitive process of fact-finders, and the relative plausibility model of this process).

¹⁶⁹ See *supra* Section 3.2 (explaining the use of likelihood ratios as way of understanding the meaning of the *weight of evidence* in a fact-finding context).

¹⁷⁰ See *supra* Section 4 (highlighting several implications of the likelihood theory of fact-finding, particularly where they contrast with those of extant theories.)

¹⁷¹ See *supra* notes 143–144 and accompanying text.

and fact-finder hostility to naked statistical evidence.¹⁷² And it answers the unaddressed question in *Peña Rodriguez*, explaining the damage caused by fact-finder bias in all its forms—not just where racially motivated—and showing just how fundamentally bias, prejudice, and prior beliefs affect the deliberative process.¹⁷³

This is not to say that the likelihood theory of fact-finding is without limitations. Like any formal model, this theory 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 operates from within a frame of reference, and so cannot itself guide the framing of a question. This complicates the process of identifying what is or isn't a prior probability in some cases,¹⁷⁴ and leaves unresolved some difficult questions about the scope of legal fact-finding.¹⁷⁵ And while the proposed likelihood theory works well to explain the assessment of deliberate actions and historic facts—the type of *facts* that legal fact-finding is most often about—it is an awkward fit to situations in which material facts are truly random variables,¹⁷⁶ or in which the fact-finder is required to predict future events.¹⁷⁷ In these settings, there may yet be room for prior beliefs and Bayesian probability analysis in legal fact-finding.¹⁷⁸

But these limitations are narrow, and the benefits of the likelihood approach are real. Beyond the unifying and explanatory advantages already mentioned, this likelihood theory of fact-finding simplifies the cognitive model of fact-finding by showing that *every* burden of persuasion can be understood in terms of the very same framework. That framework, in turn, is an intuitive and purely comparative approach of

¹⁷² See *supra* notes 145–148 and accompanying text.

¹⁷³ See *supra* notes 3–9, 155–159 and accompanying text.

¹⁷⁴ See *supra* note 144 (discussing habitual act evidence as a possible example).

¹⁷⁵ See *supra* note 134 (discussing ambiguity in the scope of fact-finding).

¹⁷⁶ See *supra* note 144 (interpreting habit evidence as a truly-random fact situation).

¹⁷⁷ Preventative injunctive remedies are an obvious and important example of this type of predictive fact-finding. See OWEN M. FISS, *THE CIVIL RIGHTS INJUNCTION* 8–12 (Indiana Univ. Press 1978) (distinguishing future-oriented *preventative* injunction from retrospective *reparative* injunctions).

¹⁷⁸ This is not to say that even in a predictive setting, probability or belief-based proof concepts are a comfortable fit to the legal context. See, e.g., Uviller, *supra* note 22, at 848 (noting, in the context of character-based prediction of conduct that “The law . . . has been particularly squeamish about predictive evidence; the result has been uncertain, inconsistent, and ill-defined rules.”).

iteratively weighing evidence between competing factual theories. The result is a theory of legal fact-finding that is both internally consistent, and consistent with common sense and actual trial practice.

To unlock the benefits of this new theory, we need to adjust the way we think about and describe fact-finding at trial. To be clear, I am not proposing that fact-finders should be instructed in the rigors of likelihood analysis. Nor am I proposing any semblance of the feared *trial by numbers* in which calculators replace pads and pencils.¹⁷⁹ And I also do not think that simply changing the language from *probability* to *likelihood* will change anything in practice. The distinction between these concepts is too esoteric for words alone to make a difference.

Rather, adjustments are needed in both our language *and* our understanding of fact-finding. Probability terms are now scattered throughout the law of evidence. The Federal Rules of Evidence define “relevant evidence” as anything having “any tendency to make a [material] fact more or less probable than it would be without the evidence.”¹⁸⁰ The Model Rules of Evidence define “finding a fact” as “determining that [the fact’s] existence is more probable than its non-existence.”¹⁸¹ Jury instructions commonly define burdens of persuasion in probability terms.¹⁸² And some scholars would go so far as to claim that “a lawsuit is essentially a search for probabilities.”¹⁸³ A predictable response is that these terms are not intended to be interpreted in a technically accurate sense, but that only solidifies the importance of understanding what the terms could, and do, mean in legal fact-finding.

As mentioned in the introduction of this paper, Richard Royall provides a helpful way to conceptualize the options:¹⁸⁴

- Bayesian 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?”

¹⁷⁹ Calculators *with* pads and pencils, I support.

¹⁸⁰ FED. R. EVID. 401.

¹⁸¹ AMERICAN LAW INSTITUTE, MODEL CODE OF EVIDENCE, Rule 1(5) (internal quotation marks omitted).

¹⁸² See *supra* note 44 and accompanying text.

¹⁸³ MCCORMICK ON EVIDENCE, *supra* note 51, at 669.

¹⁸⁴ See *supra* note 9 and accompanying text.

That is, Bayesian probability is fundamentally a description of belief; likelihood is fundamentally a description of weight of evidence.¹⁸⁵

Over the years, many scholars have written that fact-finders' beliefs are the thing we care about, not some abstract concept of the weight of the evidence.¹⁸⁶ But this article shows the exact opposite to be true. Most fact-finding only makes sense when it is approached as a question of weight-of-evidence alone.¹⁸⁷ Not only do we not care about fact-finders' personal beliefs in fact-finding, but to insure the just determination of disputed factual questions, the role of personal beliefs must be as minimized and extracted from fact-finding, as far as it possibly can be.

This, in turn, suggests a need for adjustments in how burdens of persuasion are understood by judges and explained to juries. While fact-finders should, as ever, use their experience and intuition to guide their deliberations, we should not ask for, nor even mention, fact-finders' personal beliefs about what happened. The fact-finders' task is not to gauge their personal beliefs about the facts, but instead to weigh the evidence produced at trial, comparing the relative plausibility of conflicting factual theories—the relative consistency of these theories with observed evidence—in deciding the outcome of a case.

In some respects, this simply aligns our language and theory with what the trial process already strives to do. Many fact-finders doubtless *do* already operate as the likelihood theory requires—refusing to assign liability on purely statistical evidence, for example.¹⁸⁸ Rules of evidence prohibiting character reasoning already *do* reflect likelihood theory's rejection of prior probabilities.¹⁸⁹ And the protections of *voir dire* and jury instructions probably *do* dilute the effect of prior beliefs and fact-

¹⁸⁵ Royall, *supra* note 9 at 131.

¹⁸⁶ *E.g.* James, *supra* note 111, 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 12, 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.”).

¹⁸⁷ See SUSAN HAACK, EVIDENCE MATTERS: SCIENCE, PROOF, AND TRUTH IN LAW 51–53 (2014) (expressing the concurrent view that the epistemic language of burdens of persuasion and jury instructions often reveals a primary concern with the objective strength of evidence in a case, as opposed to the subjective beliefs of the fact-finder).

¹⁸⁸ See *supra* notes 141–142, 145–148 and accompanying text.

¹⁸⁹ See *supra* notes 143–144 and accompanying text.

finder bias in many cases.¹⁹⁰ But fortuitous success is weak assurance against future mistake,¹⁹¹ and these diverse rules and practices are no substitute for clarity about the underlying theory of legal fact-finding.¹⁹² The likelihood theory of fact-finding supplies that clarity, explaining and justifying fact-finding as a guided search for the most likely story.

¹⁹⁰ See, e.g., FED. R. CIV. P. 47 (selecting jurors in federal civil trials); FED. R. CRIM. P. 24 (selecting jurors in federal criminal trials); FED. R. EVID. 105 (limiting the scope of evidence by instruction to the jury in federal trial); Seventh Circuit, Federal Civil Jury Instructions, *supra* note 10, at § 1.06 (instructing the jury on what is *not* evidence).

¹⁹¹ See Learned Hand, *Have the Bench and Bar Anything to Contribute to the Teaching of Law?*, 24 MICH. L. REV. 466, 470 (1925) (“[A] right result reached by unsound reasons gives no assurance of permanent acquisition.”).

¹⁹² Lempert, *supra* note 27, at 1056 (“Whatever enables lawyers to think more clearly is of practical importance.”).