EVIDENCE, PROBABILITY, AND THE BURDEN OF PROOF

Ronald J. Allen[*] and Alex Stein[**]

This Article analyzes the probabilistic and epistemological underpinnings of the burden of proof doctrine. We show that this doctrine is best understood as instructing factfinders to determine which of the parties' conflicting stories makes most sense in terms of coherence, consilience, causality, and evidential coverage. By applying this method, factfinders should try—and will often succeed—to establish the truth, rather than a statistical surrogate of the truth, while securing the appropriate allocation of the risk of error. Descriptively, we argue that this understanding of the doctrine—the "relative plausibility theory"—corresponds to our courts' practice. Prescriptively, we argue that the relative-plausibility method is operationally superior to factfinding that relies on mathematical probability. This method aligns with people's natural reasoning and common sense, avoids paradoxes engendered by mathematical probability, and seamlessly integrates with the rules of substantive law that guide individuals' primary conduct and determine liabilities and entitlements. We substantiate this claim by juxtaposing the extant doctrine against two recent contributions to evidence theory: Professor Louis Kaplow's proposal that the burden of proof should be modified to track the statistical distributions of harms and benefits associated with relevant primary activities; and Professor Edward Cheng's model that calls on factfinders to make their decisions by using numbers instead of words. Specifically, we demonstrate that both models suffer from serious conceptual problems and are not feasible operationally. The extant burden of proof doctrine, we conclude, works well and requires no far-reaching reforms.

TABLE OF CONTENTS

INTRODUCTION

Legal factfinding, like most real life decision-making, involves decision under uncertainty.[1] Consequently, the legal system has adopted a set of decision rules to instruct judges and jurors how to decide cases in the face of uncertainty. These rules are collectively known as the burden of proof.[2] They include the well-known requirement that all accusations against the defendant in criminal cases be proven "beyond a reasonable doubt."[3] For defenses that an otherwise guilty defendant may raise, the rules often require proof by a "preponderance of the evidence"[4] or proof by "clear and convincing evidence."[5] In civil litigation, the burden of proof tends to treat plaintiffs and defendants as equals, normally requiring each party to prove her allegations—the plaintiff's cause of action and the defendant's affirmative defenses—by a "preponderance of the evidence."[6] For allegations of crime and fraud in civil cases, the proof burden is often set to "clear and convincing evidence"—a special proof requirement that also applies in proceedings that might deny a person certain civil rights, such as deportation,

---

    1.   *See* ALEX STEIN, FOUNDATIONS OF EVIDENCE LAW 34–36 (2005) (underscoring the inevitable presence of uncertainty in adjudicative factfinding).

    2.   *See generally* CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, EVIDENCE §§ 3.1–3.3, 3.11–3.12, at 103–12, 134–42 (5th ed. 2012) (discussing civil and criminal burdens of proof).

    3.   *Id.* §§ 3.11–3.12 at 134–42.

    4.   *Id.* §§ 3.12 at 136–42.

    5.   *See, e.g.*, 18 U.S.C. § 17(b) (2006) ("The defendant has the burden of proving the defense of insanity by clear and convincing evidence.").

    6.   *See* MUELLER & KIRKPATRICK, *supra* note 2, § 3.3, at 111.

denaturalization, involuntary confinement to a mental institution, and removal of parental rights.[7]

Some of these rules are entrenched in the Constitution;[8] most are a matter of state policy. A defendant's right to be acquitted when one or more elements of the crime are not proven beyond a reasonable doubt is part of his entitlement to "due process of law" under the Fifth and Fourteenth Amendments.[9] The Due Process Clause also includes the "clear and convincing evidence" requirement for allegations that may lead to a denial of civil rights.[10] The Ex Post Facto Clause does not allow the burden of proof—in criminal cases and with regard to statutory prohibitions that are not explicitly criminal but have a punitive intent—to be altered retroactively.[11] Finally, the *Erie* doctrine (widely considered "quasi-constitutional") gives the states precedence over Congress in setting up burdens of proof for diversity suits.[12]

Legal scholars have long recognized the centrality of the burden of proof and its effects on individuals' entitlements and primary activities.[13] This recognition led scholars to investigate the conceptual foundations of the burden of proof, as well as how it integrates into the factfinding process as a whole. Economically minded scholars have investigated the connections between the burden of proof, risk of error, primary behavior, and cost of litigation.[14] Moral theorists, beginning with Immanuel Kant, have tried to identify the evidentiary

---

7. *Id.* § 3.3, at 112.

8. *See* Alex Stein, *Constitutional Evidence Law*, 61 VAND. L. REV. 65, 79–82 (2008) (attesting that the "proof beyond a reasonable doubt" requirement for criminal convictions and the "clear and convincing evidence" standard for allegations that justify deprivations of civil rights and liberties are mandated by due process).

9. *Id.* at 79–80.

10. *Id.* at 81–82.

11. *Id.* at 99–101.

12. *Id.* at 98–99.

13. *See, e.g., Symposium on Presumptions and Burdens of Proof*, 17 HARV. J. L. & PUB. POL'Y 613 (1994).

14. *See* Bruce L. Hay & Kathryn E. Spier, *Burdens of Proof in Civil Litigation: An Economic Perspective*, 26 J. LEGAL STUD. 413, 418–21 (1997) (analyzing burden of proof as an instrument for reducing the cost of litigation); Gideon Parchomovsky & Alex Stein, *The Distortionary Effect of Evidence on Primary Behavior*, 124 HARV. L. REV. 518, 530–42 (2010) (explaining people's primary behavior as motivated by the burdens of proof and other evidentiary requirements); Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1502–07 (1999) (unfolding economic analysis of the burden of proof as a tool for reducing the cost of errors and error-avoidance as a total sum); David Rosenberg, *The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System*, 97 HARV. L. REV. 849, 861–67 (1984) (carrying out economic analysis of the burden of proof and identifying the limits of the "preponderance" standard in tort cases with uncertain causation); Chris W. Sanchirico, *Games, Information and Evidence Production: With Application to English Legal History*, 2 AM. L. & ECON. REV. 342, 343–44 (2000) (unfolding an account of proof burdens that uses evidence production as a proxy for determining the harmfulness of primary behavior); Chris W. Sanchirico, *Relying on the Information of Interested—and Potentially Dishonest—Parties*, 3 AM. L. & ECON. REV. 320 (2001) (analyzing the proof burdens' effect on primary behavior).

minimum that could justify an imposition of punishment or other deprivation on a person who may not have committed the alleged wrong.[15] The body of literature produced by these scholars is rich, insightful, and multifaceted.

This Article investigates the relationship between evidence, probability, and the burden of proof. We examine what factfinders do when they decide cases by applying the controlling proof burden. We demonstrate that factfinders decide cases predominantly by applying the relative plausibility criterion guided by inference to the best explanation, rather than by using mathematical probability.[16] Indeed, we show that our courts apply mathematical probability only to a small number of well-defined categories of cases.[17] We then evaluate this practice and commend it on the grounds of both pragmatism and principle.

We show that the relative plausibility approach outperforms mathematical probability operationally and normatively. Application of mathematical probability in the courts of law engenders paradoxes and anomalies that are not easy to avoid or explain away. Relative plausibility, on the other hand, faces no such predicaments. A further advantage is its alignment with the natural reasoning of ordinary people, which reduces the cost of adjudication and helps the legal system guide individuals' behavior. Last, but not least, relative plausibility is the best available tool to get factfinders to the actual facts of the case they are asked to resolve. Mathematical probability, on the other hand, abstracts away from those facts. As a substitute, it prods factfinders to derive their decisions from the general frequencies of events.

We combine this discussion with our critique of the two most recent contributions to the burden of proof literature: Louis Kaplow's radical proposal to revamp the burden of proof doctrine[18] and Edward Cheng's introduction of a new mathematical tool for factfinders' use.[19]

Kaplow proposes a complete overhaul of the burden of proof doctrine, which he criticizes for having "almost nothing to do with what matters for society."[20] His analysis starts from the fundamental premise that, because certainty in factfinding is not within the legal system's reach, the system should strive to achieve a socially optimal distribution of adjudicative errors: mistaken impositions

---

15.   *See* Ernest J. Weinrib, *Private Law and Public Right*, 61 U. TORONTO L.J. 191, 210 (2011) (explaining Kant's rationalization of the burden of proof as "an aspect of the defendant's innate right to be considered beyond reproach in the absence of an act that wrongs another").

16.   For foundational articles on this subject, see Ronald J. Allen, *A Reconceptualization of Civil Trials*, 66 B.U. L. REV. 401, 403 (1986); Ronald J. Allen, *Factual Ambiguity and a Theory of Evidence*, 88 NW. U. L. REV. 604 (1994) [hereinafter Allen, *Factual Ambiguity*]; Ronald J. Allen, *The Nature of Juridical Proof*, 13 CARDOZO L. REV. 373 (1991).

17.   *See infra* note 108 and accompanying text.

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

19.   Edward K. Cheng, *Reconceptualizing the Burden of Proof*, 122 YALE L.J. 1254, 1258–59 (2013).

20.   Kaplow, *supra* note 18, at 789.

of legal liability ("errors of commission") and mistaken failures to impose legal liability ("errors of omission"). According to Kaplow, optimal distribution of those errors does not correlate with the extent to which courts' decisions are accurate. As established in Kaplow's previous work, accuracy *ex post* has no value in and of itself.[21] Distribution of adjudicative errors—regardless of the accuracy rate it produces over a run of cases—thus ought to promote a different goal: It ought to incentivize *ex ante* socially optimal primary behavior.

Consistent with this vision, Kaplow criticizes the burdens of persuasion that function as proof requirements under extant law: "preponderance," "beyond a reasonable doubt," and "clear and convincing evidence."[22] These probability standards, Kaplow argues, work to achieve accuracy *ex post*—an economically inefficient goal that our legal system ought to abandon.[23] They ought to be replaced by a different legal mechanism that incentivizes socially desirable conduct *ex ante*.

To implement his idea, Kaplow argues for the creation of what he calls "evidence thresholds."[24] This novel mechanism is the core insight of Kaplow's normative theory. Evidence that goes into Kaplow's thresholds informs courts about the effects of the relevant activity—harmful and socially useful, or "benign"[25]—across a series of cases. This evidence will associate different activities with different concentrations of harm and benefit. Some of those concentrations yield a negative tradeoff; others do not. Policymakers consequently will desire to suppress activities associated with the undesirable concentrations of harm versus benefit, while allowing other activities to take place. Policymakers can achieve this result by setting up rules that sanction the undesirable concentrations of harm versus benefit. Sanctions will follow according to a sliding scale of the probability in which the higher the predominance of harm in the mix, the lower the probability needed for liability; and conversely, the lower the risk of harm, the higher the probability needed. According to Kaplow, this myriad of rules should replace the conventional burden of proof doctrine.[26]

Edward Cheng recasts the burden of proof doctrine in terms of standard mathematical probability.[27] Kaplow's theory presupposes that the extant proof requirements—"preponderance," "beyond a reasonable doubt," and "clear and convincing"—have numerical equivalents on the probability scale between 0 and 1 and that courts associate these requirements with mathematical probability. Cheng does not accept this presupposition, and for a good reason: Courts generally do not

---

21.   *See* Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307 (1994).

22.   Kaplow, *supra* note 18, at 742–44.

23.   *Id.* at 784–89.

24.   *Id.* at 756–62.

25.   Kaplow uses the term "benign" and the awkward term "benignancy," for which we substitute the more straightforward term "benefit" and its derivatives.

26.   Kaplow, *supra* note 18, at 755–72.

27.   Cheng, *supra* note 19, at 1259–65.

use mathematical probability in applying the burden of proof doctrine.[28] Importantly, the prevalent academic opinion approves this practice: Most evidence scholars believe that adjudicative factfinding is fundamentally incompatible with mathematical probability.[29] Mathematical probability sometimes allows policymakers to evaluate the overall performance of a rule or a set of rules and macromanage the legal system as a whole.[30] Carrying this tool to the process of determining individual facts is broadly considered a bad idea.[31] Cheng's article undertakes to overturn this widely accepted "incompatibility thesis."[32]

To discharge this task, Cheng develops a mathematical method that removes the problems that make "trial by mathematics" operationally nonfeasible and normatively unattractive.[33] One of those problems—the most difficult one, in the eyes of many—is the "conjunction paradox."[34] Consider a breach-of-contract suit that needs to be proven by a preponderance of the evidence, denoted as a mathematical probability greater than 0.5. Assume that the plaintiff makes two mutually independent allegations: (1) The defendant and she contracted for delivery of goods and (2) the defendant breached the contract by not delivering the goods that he undertook to deliver. Assume further that the evidence the parties adduce indicates that each of these allegations has a 0.7 probability. The conventional understanding of the burden of proof doctrine holds that the court

---

28.   *See* STEIN, *supra* note 1, at 238–39.

29.   *See* William L. Twining & Alex Stein, *Introduction* to EVIDENCE AND PROOF *in* VOL. XI OF INTERNATIONAL LIBRARY OF ESSAYS IN LAW AND LEGAL THEORY xxi–xxiv (William L. Twining & Alex Stein, eds. 1992) (discussing the probability debate and underscoring the mismatch between mathematical probability and adjudicative factfinding); Symposium, BAYESIANISM AND JURIDICAL PROOF, in 1 INT. J. EVIDENCE & PROOF 253, 254– 360 (Ron Allen & Mike Redmayne eds., 1997) (debating the applicability of mathematical probability to adjudicative factfinding).

30.   *See, e.g.*, Alex Stein, *Inefficient Evidence* 1 (Benjamin N. Cardozo School of Law, Cardozo Legal Studies Faculty Research Paper No. 380, 2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2199601   (using   mathematical probability to explain and guide the legal system's macromanagement of evidence).

31.   *See, e.g.*, L. JONATHAN COHEN, THE PROBABLE AND THE PROVABLE (1977) (unfolding a broad philosophical theory that identifies a fundamental misfit between mathematical probability and adjudicative factfinding); Ronald J. Allen, *Rationality, Algorithms, and Juridical Proof: A Preliminary Inquiry*, 1 INT. J. EVIDENCE & PROOF 254, 275 (1997) (specifying incompatibilities between mathematical probability and juridical proof, while underscoring the virtues of natural reasoning to the best explanation); Craig R. Callen, *Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law*, 57 IND. L.J. 1, 2–3 (1982) (demonstrating that application of mathematical probability in courts of law requires factfinders to carry out unbearably complex calculus); Alex Stein, *Judicial Fact-Finding and the Bayesian Method: The Case for Deeper Scepticism about their Combination*, 1 INT'L. J. EVIDENCE & PROOF 25, 41 (1996) (demonstrating that the Bayesian approach to adjudicative factfinding that employs subjective probabilities is tautological).

32.   *See* Cheng, *supra* note 19, at 1259–65.

33.   *Id.* at 1258–62.

34.   *Id.* at 1263–65. *See also* COHEN, *supra* note 31, at 58–61 (original statement of the conjunction paradox in evidence law).

2013]

should rule in favor of the plaintiff, whose case is much stronger than the defendant's. Under the conventional understanding of probability, however, the plaintiff's case is actually weaker than the defendant's. The combined probability of the plaintiff's allegations against the defendant is 0.49 (0.7 × 0.7)—just below the preponderance threshold. The probability of the defendant's claim "I made no contract with the plaintiff or, alternatively, committed no breach" is 0.51 (0.3 + 0.3 − 0.3²). Hence, the defendant should prevail.

This mathematical outcome contradicts legal doctrine and common sense, which is why it received the name "the conjunction paradox."[35] Evidence scholars, including us, have tried to resolve this paradox or somehow explain it away.[36] Cheng's article makes an important addition to these efforts by developing a novel method to avoid the paradox. This method shifts away from a categorical assessment of probability to a comparative assessment. If successful, it would refute the incompatibility thesis and vindicate trial by mathematics.

Cheng argues that the preponderance requirement (along with all other probability thresholds incorporated in the burdens of proof) should be understood in comparative, rather than categorical, terms.[37] Courts should compare the individual probabilities attaching to the plaintiff's factual allegations and to the defendant's story. This comparison will determine whose case is stronger. As Cheng explains, courts should proceed in the same way in which scientists choose between competing hypotheses.[38] This decision-making framework will not allow the defendant in our breach of contract case to rely on the probability of the disjunctive scenario "I made no contract with the plaintiff, and if I did make it somehow, I did not breach it." This scenario is counterfactual and hence does not form a hypothesis comparable with the plaintiff's allegations of *fact*. The probabilities of the parties' comparable *factual* allegations thus show 0.7 on the plaintiff's side and 0.3 on the defendant's side. This mathematical outcome aligns with the decision that factfinders would reach by applying the relative plausibility method.[39] Cheng's probabilistic account thus connects a mathematical approach to factfinding to the best present understanding of burdens of persuasion.[40]

Our critique of Kaplow's theory is threefold. First, we show that his proposal cannot be adopted because of its enormous (essentially, infinite) informational costs. Second, Kaplow's evidence thresholds are direct analogues of

---

35.   *See* STEIN, *supra* note 1, at 49–50.

36.   *See Id.* at 49–56; Ronald J. Allen & Sarah A. Jehl, *Burdens of Persuasion in Civil Cases: Algorithms v. Explanations*, 2003 MICH. ST. L. REV. 893, 944; 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 (1998) [hereinafter Stein, *Uncertainty and Fact-Finding*]; Alex Stein, *Of Two Wrongs that Make a Right: Two Paradoxes of the Evidence Law and their Combined Economic Justification*, 79 TEX. L. REV. 1199, 1199–2000 (2001) [hereinafter Stein, *Two Wrongs*]; Ronald J. Allen, *Book Review: Laudan, Stein, and the Limits of Theorizing About Juridical Proof*, 29 L. & PHIL. 195, 225–26 (2010).

37.   Cheng, *supra* note 19, at 1259–61.

38.   *Id.* at 1257, 1276–77.

39.   *Id.* at 1259–62.

40.   *Id.* at 1259–65.

Bayesian likelihood ratios.[41] Bayes' Theorem shows that basing decisions upon likelihood ratios instead of the posterior probabilities that account for all relevant information is a mistake.[42] Under the Bayesian framework, the optimal proof burden in any given context will derive from the desired ratio of false positives ("errors of commission") and false negatives ("errors of omission"), although the formulation of that ratio is, as we discuss, complicated—much more so than Kaplow seems to realize. This formulation of the burden of proof explains and to a significant extent justifies the conventional view.

     Last, the conventional proof burdens track the substantive definitions of tort and criminal liability that require courts to base liability decisions on the actor's *ex ante* information, but Kaplow paid no attention to those definitions. This omission has two implications. First, substantive definitions of liability—both civil and criminal—go far toward aligning courts' applications of the conventional burdens of proof with the *ex ante* distributions of harm versus benefit. We show that taking this factor into consideration substantially vindicates the conventional approach to the burden of proof. The conventional burden of proof doctrine is more sophisticated and better aligned with efficiency than Kaplow believes it to be. Similarly, Kaplow's theory abstractly categorizes individuals' activities as harmful and beneficial without regard to the specific nature of the primary behavior. As a result, the theory does not distinguish between accidents, contract breaches, and crimes. The theory's failure to address these harms separately misses an important—indeed pivotal—characteristic of our legal system. The system prescribes separate combinations of proof burdens and other rules for accidents, breaches of contract, and crimes. For liability flowing from accidents, the system constructs evidentiary rules that motivate prospective wrongdoers to base their conduct on the *ex ante* probability of causing harm. These rules include liability presumptions driven by regulatory statutes and probability-based recovery of tort compensation. Accident law thus may not need Kaplow's evidence thresholds. Contracts plainly require no such thresholds either, as parties are generally best situated to design their own evidentiary mechanisms for resolving allegations of breach,[43] which both substantive and procedural laws unequivocally permit.[44] The conventional burden of proof functions in contract law as a mere default,[45] which Kaplow does not (and cannot) criticize.

---

     41.   *See infra* Part II.B. Kaplow's illustrations of how evidence thresholds are supposed to work strengthen this association. Kaplow, *supra* note 18, at 785–86.
     42.   For an explanation of Bayes' Theorem, *see* Alex Stein, *The Flawed Probabilistic Foundation of Law and Economics*, 105 Nw. U. L. Rev. 199, 211–12 (2011), and sources cited therein.
     43.   *See* Robert E. Scott & George G. Triantis, *Anticipating Litigation in Contract Design*, 115 Yale L.J. 814, 814 (2006).
     44.   *See* Robert G. Bone, *Party Rulemaking: Making Procedural Rules Through Party Choice*, 90 Tex. L. Rev. 1329, 1330 (2012); John W. Strong, *Consensual Modifications of the Rules of Evidence: The Limits of Party Autonomy in an Adversary System*, 80 Neb. L. Rev. 159, 160 (2001).
     45.   *See* Stein, *Uncertainty and Fact-Finding*, *supra* note 36, at 341–44.

Another example of the consequences of failing to attend to different forms of liability involves the criminal law. Criminal law aspires to optimal deterrence by adjusting applicable penalties while requiring the prosecution's evidence to establish a very high posterior probability of guilt.[46] This is a very sensible way to reduce crime while protecting innocents from wrongful conviction. As Gary Becker demonstrated long ago, penalty adjustments can achieve optimal deterrence more expediently and cost-effectively than adjustments of law enforcement.[47] The reason is obvious. Enforcement efforts that require information are expensive: Indeed, Kaplow acknowledges that setting up his evidentiary thresholds is a costly exercise.[48] Criminal penalties, on the other hand, can be set with a strike of a pen.

Our critique of Cheng's theory is straightforward. Cheng's theory succeeds in developing a mathematical conceptualization of the burden of proof that avoids the conjunction paradox as it is presently understood. But that is all that it does. Critically, it ignores the consequences of the systematic suppression of the probabilities of opposite scenarios. Cheng also fails to explain why it would be good for society if our courts were to use his conceptualization instead of the conventional one. We address that very question in the pages ahead. We show that the conventional proof burden, conceptualized as inference to the best explanation, does a better job in promoting the fairness and efficiency of our legal system. Unlike the mathematical understanding of the proof burden, this conceptualization gives rise to no anomalies and paradoxes.

Structurally, this Article unfolds as follows. In Part I, we explain how the conventional burden of proof doctrine works and how it promotes efficiency and fairness. In Parts II and III, respectively, we analyze and criticize Kaplow's and Cheng's theories of the burden of proof. A short Conclusion ensues.

I. THE NATURE OF THE BURDEN OF PROOF

Burdens of proof easily bear a probabilistic interpretation. In civil cases, the standard instruction tells jurors that each element of a claim and of an affirmative defense must be established by a preponderance of the evidence, where "preponderance" means more likely than not.[49] This formulation of the proof burden leads directly to the probabilistic interpretation of greater than a 0.5 probability.[50] In criminal cases, the "beyond a reasonable doubt" instruction decidedly avoids asking jurors to quantify their doubts concerning the defendant's guilt. Asking jurors to do so is tantamount to asking them to sacrifice a number of innocents in order to allow the criminal justice system to convict and punish a

---

46. *See, e.g.*, Richard A. Bierschbach & Alex Stein, *Mediating Rules in Criminal Law*, 93 VA. L. REV. 1197, 1210–12 (2007) (explaining and citing literature as to what optimal deterrence in criminal law requires).

47. *See* Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169, 180–84 (1968).

48. *See* Kaplow, *supra* note 18, at 771, 786–89.

49. *See* MUELLER & KIRKPATRICK, *supra* note 2, § 3.3, at 111–12.

50. *See, e.g.*, Allen & Jehl, *supra* note 36, at 894–95.

sufficient number of guilty offenders.[51] Despite this operational difficulty, mathematical probability can give meaning to the criminal proof burden as well; the same is true for "clear and convincing" evidence.[52] The significant questions here are whether any of these reconceptualizations are empirically accurate or normatively attractive as a potential improvement of our legal system. Our answer to both questions is no.

      Scholars' attempts at mathematizing the burden of proof follow a frequentist interpretation of probability,[53] and for good reason. Other interpretations of the concept of "probability"—logical, propensity, and subjective beliefs[54]—make no sense at all in the juridical context.[55] The frequentist account of probability, however, does not do much better. Courts resort to frequentist probability in some very specific contexts.[56] Outside these contexts, frequentist probability is of no use. Pragmatism and substance drive our courts' general rejection of this probability.[57] Courts have no information about the relative

---

51.   *See generally* Alexander Volokh, *n Guilty Men*, 146 U. PA. L. REV. 173, 198 (1997).

52.   Probability thresholds for these burdens can be set at any appropriate level, for example: 0.95 ("beyond a reasonable doubt") and 0.75 ("clear and convincing evidence").

53.   Frequentist probability is a system of reasoning that associates an event's chances of occurring with instantial multiplicity. Under this system, an event's chances of occurring are favorable when it falls into the majority of the observed events. Conversely, an event's chances of occurring are not favorable when it falls into the minority of the observed events. An event's probability consequently equals the number of cases in which it occurred divided by the totality of relevant cases. *See* L. JONATHAN COHEN, AN INTRODUCTION TO THE PHILOSOPHY OF INDUCTION AND PROBABILITY 47–48 (1989); *see also* STEIN, *supra* note 1, at 143–48 (discussing mathematical approaches to the burden of proof and their uniform reliance on frequentist probability).

54.   *See generally* DONALD GILLIES, PHILOSOPHICAL THEORIES OF PROBABILITY 1 (2000) (explaining different versions of probability); COHEN, *supra* note 53, at 53–80 (analyzing logical, propensity-based, and subjectivist interpretations of "probability" and explaining their limitations).

55.   *See* Michael S. Pardo & Ronald J. Allen, *Juridical Proof and the Best Explanation*, 27 L. & PHIL. 223, 227–38 (2008); Alex Stein, *Bayesioskepticism Justified*, 1 INT. J. EVIDENCE & PROOF 339, 342 (1997) (formal demonstration of circularity and self-reference that plague the subjectivist version of probability as applied in juridical context); Stein, *supra* note 31, at 41 (rejecting the subjective-belief version of probability as tautological).

56.   *See infra* note 108 and accompanying text.

57.   *See* United States v. Shonubi, 998 F.2d 84 (2d Cir. 1993) (reversing a lower court decision in United States v. Shonubi, 802 F. Supp. 859, 860–64 (E.D.N.Y. 1992), that used mathematical probability to determine a fact aggravating the defendant's crime and sentence); Stein, *Two Wrongs*, *supra* note 36, at 1204 n.6, 1205 (citing different jury instructions that run contrary to mathematical probability); *see also* Ronald J. Allen & Michael S. Pardo, *The Problematic Value of Mathematical Models of Evidence*, 36 J. LEGAL STUD. 107, 130–35 (2007) (rationalizing the Second Circuit's reversal of the trial judge's decision in *Shonubi* by the judge's failure to carve out the relevant reference class). *Cf.* United States v. Veysey, 334 F.3d 600, 604–06 (7th Cir. 2003), *cert. denied* 540 U.S. 1129 (2004) (approving defendant's arson conviction based on actuarial testimony estimating that

frequencies of relevant events. Equally important, as we demonstrate below, our courts have a strong substantive preference for the epistemic mode of factfinding: a system of reasoning guided by inference to the best explanation. In what follows, we describe this mode of factfinding and explain its merits.[58]

A. Adjudicative Factfinding as Inference to the Best Explanation

We begin with a simple, but oft-neglected, observation: The coin of the legal realm is truth. Factfinders (judges or jurors) operating in that realm try to reconstruct an event that involves the parties to the trial. They focus on the specific occurrences in which the parties (or a single party when that party is a criminal defendant) took part. As part of this reconstruction process, factfinders receive evidence from the parties and evaluate the significance of that evidence through their collective experience about how the world works. Factfinders then juxtapose the parties' conflicting accounts of the event and ask themselves which of those accounts makes the most sense. By applying this method, factfinders try to get to the truth itself, rather than to a statistical surrogate of the truth. They do not base their decisions on the frequencies of events that resemble the event they are trying to reconstruct (even when those frequencies are available). Instead, they rely upon case-specific evidence that uncovers the details and individual characteristics of the event in question.[59]

In the absence of reliable relative frequencies, there is only one other manner to operationalize a frequentist account of burdens of persuasion. The civil proof burden can be understood as requiring a plaintiff to show that, of all the ways the universe might have been on the day in question, half plus one favor liability. Similarly, if this frequentist mode of proof were to apply in criminal cases, a defendant might be able to bring in the phone book to show that many people—potentially, millions—could have committed the alleged crime, and the prosecution would have to establish that they did not. Our legal system has not adopted this understanding of the proof burden because it would be virtually (if not altogether) impossible for plaintiffs and prosecutors to ever win. The burden of proof doctrine, as applied by our courts, took a different path that aligns with common sense: It endorsed the relative plausibility criterion for factual findings.

Relative plausibility takes hold at the very beginning of a case. Litigation starts off with opening statements as to how the world was the day in question. At the end of the case, each side attempts to close the deal by weaving together a

---

the chances of four residential fires occurring by accident during the relevant period were 1 in 1.773 trillion); STEIN, *supra* note 1, at 205–07 (criticizing the arson conviction in *Veysey* for failure to align with the "principle of maximal individualization").

    58.   One of the present Authors (Alex Stein) is more sanguine about giving normative prescriptions than the other (Ronald Allen). This Article's normative claims will therefore be parsimonious: In what follows, we compare the workings of the conventional factfinding method with the consequences flowing from the adoption Cheng's and Kaplow's reforms.

    59.   *Cf.* STEIN, *supra* note 1, at 91–106 (introducing the "principle of maximal individualization").

coherent narrative in the closing argument, where plausibility is determined in common sense terms.[60] To win the plausibility contest, evidence that a party relies upon must unfold a narrative that makes sense to a natural reasoner: a layperson. There is no algorithm for "plausibility;" the variables that inform judgments of plausibility are all the things that convince people that some story may be true, including coherence, consistency, coverage of the evidence, completeness, causal articulation, simplicity, and consilience (understood as the breadth of the explanation).[61] Factfinders then 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.[62] Theoretically, a defendant can simply deny the plaintiff's complaint, which is precisely what would be the case were the probabilistic account of the proof burden descriptively accurate, but this virtually never occurs.[63] Indeed, even in criminal cases, a defendant must offer a factual alternative to the story the prosecution tells or face a heightened risk of conviction: If a plausible story of guilt is on the table and there is no alternative, well, "the dog did not bark . . . ."[64] The data is striking that without barking dogs, there is a high probability of conviction.[65] This consequence for parties with

---

60.   For an excellent recent account of this method, *see* Lisa Kern Griffin, *Narrative, Truth, and Trial*, 101 GEO. L.J. 281, 293 (2013); *see also* Luke Meier, *Probability, Confidence, and* Twombly*'s Plausibility Standard* (unpublished manuscript, May 2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2271802 (recommending plausibility as a standard for directed-verdict decisions).

61.   *See* Pardo & Allen, *supra* note 55, at 227–36.

62.   *See* Shari Seidman Diamond et al., *Juror Questions During Trial: A Window Into Juror Thinking*, 59 VAND. L. REV. 1927, 1934 (2006); Nancy Pennington & Reid Hastie, *The Story Model for Juror Decision Making*, *in* INSIDE THE JUROR: THE PSYCHOLOGY OF JUROR DECISION MAKING 192, 194–99 (Reid Hastie ed., 1993).

63.   *See* Shari Seidman Diamond et al, *Revisiting the Unanimity Requirement: The Behavior of the Non-Unanimous Civil Jury*, 100 NW. U. L. REV. 201, 212 (2006) ("The deliberations of these 50 cases revealed that jurors actively engaged in debate as they discussed the evidence and arrived at their verdicts. Consistent with the widely accepted 'story model,' the jurors attempted to construct plausible accounts of the events that led to the plaintiff's suit. They evaluated competing accounts and considered alternative explanations for outcomes."). For comprehensive psychological studies verifying the prevalence of story-based factfinding in our courts, see W. LANCE BENNETT & MARTHA S. FELDMAN, RECONSTRUCTING REALITY IN THE COURTROOM: JUSTICE AND JUDGMENT IN AMERICAN CULTURE 3 (1981); Reid Hastie & Nancy Pennington, *Explanation-Based Decision Making*, *in* JUDGMENT AND DECISION MAKING: AN INTERDISCIPLINARY READER 212, 212–28 (Terry Connolly et al, eds., 2d. ed., 2000).

64.   *Cf.* Sir Arthur Conan Doyle, *Silver Blaze*, in THE COMPLETE SHERLOCK HOLMES 383, 397 (1953) (famous detective character, Sherlock Holmes, drawing a crucial inference from the fact that "[t]he dog did nothing in the night-time").

65.   *See* Larry Laudan & Ronald J. Allen, *The Devastating Impact of Prior Crimes Evidence and Other Myths of the Criminal Justice Process*, 101 J. CRIM. L. & CRIMINOLOGY 493, 504–06 (2011).

2013] BURDEN OF PROOF unexplained denials of opponents' allegations has been identified in the literature and named a "tactical burden of proof."[66]

The view of factfinding as a resolution of the parties' contest over the relative plausibility of evidenced stories is deeply entrenched in our legal system. A number of core evidence rules are geared toward facilitating parties' presentation of coherent narratives at trial. These rules adopt a broad concept of relevancy that renders admissible any evidence that fits into the parties' conflicting accounts of the events.[67] The rules also include special provisions for otherwise inadmissible character and hearsay evidence.[68] Under those provisions, hearsay and character evidence become admissible when they constitute an integral part of a consequential narrative that a party wants to develop.[69] Admissibility rules embrace the relative plausibility approach in a multiplicity of ways, with the burden of proof doctrine and, arguably, the relevancy doctrine as well,[70] being the only manifestations of a probabilistic approach to factfinding.[71] Unsurprisingly, the relative plausibility approach has a strong and growing presence in case law as well.[72]

A striking example of the grip that relative plausibility has on the factfinding process is the old res gestae doctrine that is still followed in many jurisdictions. This doctrine secures the admission of evidence that otherwise could not be admitted to fill in the narrative gaps in the evidence.[73] The Supreme Court placed its imprimatur on this approach in a landmark decision, *Old Chief v. United States*.[74] *Old Chief* involved an accusation that the defendant violated 18 U.S.C. § 922(g)(1)—a federal statute that prohibits possession of a firearm by

---

66. *See* Edward W. Cleary, *Presuming and Pleading: An Essay on Juristic Immaturity*, 12 STAN. L. REV. 5, 26 (1959) (explaining tactical burden of proof); *see also* Disa Sim, *Burden of Proof in Undue Influence: Common Law and Codes on Collision Course*, 7 INT. J. EVIDENCE & PROOF 221, 228–231 (2003) (modern application of the tactical proof burden).

67. *See* FED. R. EVID. 401 (defining as relevant evidence having "any tendency" to prove a fact in issue).

68. *See* Ronald J. Allen & Brian Leiter, *Naturalized Epistemology and the Law of Evidence*, 87 VA. L. REV. 1491, 1534–35 (2001).

69. *Id.*

70. *See* Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1022 (1977) (conceptualizing relevancy in terms of mathematical probability).

71. Importantly, relevancy can be understood in probabilistic terms, but it can also be understood in plausibility terms.

72. *See, e.g.*, Makor Issues & Rights, Ltd. v. Tellabs, Inc., 513 F.3d 702, 711 (7th Cir. 2008) ("The plausibility of an explanation depends on the plausibility of the alternative explanations."); United States v. Beard, 354 F.3d 691, 693 (7th Cir. 2004) ("Relative to the alternatives, the government's case was more powerful than it would have seemed in the abstract."); United States v. Newell, 239 F.3d 917, 920 (7th Cir. 2001) (using relative-plausibility mode of reasoning); Swajian v. Gen. Motors Corp., 916 F.2d 31, 34 (1st Cir. 1990) (same); MCI Commc'ns Corp. v. Am. Tel. & Tel. Co., 708 F.2d 1081, 1174 (7th Cir. 1983) (same).

73. *See* Allen & Leiter, *supra* note 68, at 1535.

74. 519 U.S. 172 (1997).

anyone with a prior felony conviction. The defendant admitted being a convicted felon, but denied the alleged possession of a firearm.[75] He offered a stipulation that he had previously been convicted of a felony offense.[76] The prosecution insisted on offering this conviction into evidence in order to make the jury aware of its specifics.[77] These specifics were analogous to the violent crime accusation that accompanied the unlawful possession charge.[78] The prosecution did not accuse the defendant of merely possessing a firearm, but also of using that firearm against the alleged victim.[79] The similarity between this accusation and the defendant's prior crime led to fears that the jury would misinterpret the prior crime as showing the defendant's propensity to violence or even modus operandi, which could prejudice the defendant in his current trial.[80] To fend off this risk, the Supreme Court held that, because the defendant's stipulation was limited to his "convicted felon" status, the trial court should have accepted his stipulation.[81]

The Court, however, also went out of its way to articulate that it was making an exceptional ruling for a case it considered exceptional.[82] The Court emphasized that normally, a defendant's admission or offer to stipulate will not prevent the prosecution from presenting evidence that unfolds its story of the crime as a natural and uninterrupted sequence of events. As the Court put it:

> [T]he accepted rule that the prosecution is entitled to prove its case free from any defendant's option to stipulate the evidence away rests on good sense. A syllogism is not a story, and a naked proposition in a courtroom may be no match for the robust evidence that would be used to prove it. People who hear a story interrupted by gaps of abstraction may be puzzled at the missing chapters, and jurors asked to rest a momentous decision on the story's truth can feel put upon at being asked to take responsibility knowing that more could be said than they have heard. A convincing tale can be told with economy, but when economy becomes a break in the natural sequence of narrative evidence, an assurance that the missing link is really there is never more than second best.[83]

The adversarial format of the American trial[84] is geared toward the same goal. This format makes parties responsible for investigating, constructing, and evidencing their competing factual narratives. Indeed, under the American system,

---

75. *Id.* at 175–76.
76. *Id.*
77. *Id.* at 177.
78. *Id.* at 174, 180–81.
79. *Id.* at 174.
80. *Id.* at 184–85.
81. *Id.* at 191–92.
82. The Court underscored that the prosecution's entitlement to present an uninterrupted narrative should be set aside because the defendant's prior conviction "would not be admissible for any purpose beyond proving status." *Id.* at 190.
83. *Id.* at 189.
84. *See, e.g.*, ROBERT A. KAGAN, ADVERSARIAL LEGALISM: THE AMERICAN WAY OF LAW 3 (2001).

each party pays her attorney's fee and cannot shift it to her opponent even when she wins the case.[85] These rules give the person best positioned to evaluate the prospect of her suit or defense the responsibility to determine her investment in the case and the right to develop and present her narrative. The rules also make each party's evidentiary task easy to understand, although, at times, difficult to discharge.

To sum up, the relative plausibility mode of factfinding involving a rigorous comparison between the parties' stories about the individual event is the norm in American courtrooms.

B. Justifying the Conventional Burden of Proof

The relative plausibility structure of adjudicative factfinding has three striking advantages. First, it solves the conjunction and all other paradoxes encountered by the frequentist account of juridical proof. Second, it aligns with ordinary people's natural reasoning. Last and most important, it focuses on the individual facts of the case and maximizes the factfinders' ability to ascertain those facts. We explain these advantages sequentially in Sections One, Two, and Three.

1. Two Modes of Factfinding

Factfinding under uncertainty is a profoundly complex endeavor that has generated much philosophical and decision theoretical disputation.[86] Adjudication further compounds the complexity by presenting difficult questions about the mode of reasoning that courts should follow. The available modes of factfinding are best described as "gambling on the truth" and "epistemic contest," but philosophers of probability have proposed other names as well.[87]

The gambling mode uses mathematical probability, a system that positions all possible scenarios, conceptualized as chances, on a scale between zero and one.[88] On that scale, zero denotes factual impossibility, while one indicates absolute certainty. Scenarios that are neither impossible nor absolutely certain are ranged between these two extremes. The probability of any such scenario is greater than zero, but less than one. The fraction of cases in which such a scenario unfolds, relative to the totality of all possible events, can consequently be represented as $1/p$. To calculate the probability of a compound event in which the scenario unfolds in conjunction with another scenario, the decision-maker needs to extrapolate the fraction of cases featuring this second scenario from the totality of events. Assume that this fraction equals $1/q$. This information allows the decision-maker to calculate the sub-fraction of cases in which the second scenario

---

85. *See* Alyeska Pipeline Serv. Co. v. Wilderness Soc'y, 421 U.S. 240, 247 (1975) ("In the United States, the prevailing litigant is ordinarily not entitled to collect a reasonable attorneys' fee from the loser.").

86. *See, e.g.*, PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE: THE USES AND LIMITS OF BAYESIANISM (Peter Tillers & Eric D. Green eds., 1988).

87. COHEN, *supra* note 53, at 4–27 (identifying the two modes of probabilistic reasoning as "Pascalian" and "Baconian").

88. *Id.* at 17–18, 56–57.

materializes within the fraction of cases represented by $1/p$, in which the first scenario also materializes. This sub-fraction equals $1/pq$.

Hence, the conjunctive probability of any two mutually independent events, A and B, equals the multiplication of A's and B's individual probabilities: $P(A\&B) = P(A) \times P(B)$ (the "product rule"). If A and B are not mutually independent in the sense that one of those events may occur in combination with the other event, then $P(A\&B) = P(A) \times P(B|A)$. Under this formulation of the product rule, $P(B|A)$ represents cases featuring event B given the presence of event A. The fraction of cases in which an event occurs always equals 1 minus the fraction of cases in which the event does not occur: $P(A) = 1 - P(not\text{-}A)$. Consequently, if A and B are two mutually exclusive events, then $P(A) = 1 - P(B)$.

These basic rules set up a framework for dealing with conditional probabilities, known as Bayes' Theorem.[89] The theorem uses the individual probabilities of two events, say, E and H, and the probability of E's occurrence in the presence of H, to calculate the probability of H's occurrence in the presence of E.

Under the product rule, $P(E\&H) = P(E) \times P(H|E)$. The same probability, restated inversely as $P(H\&E)$, equals $P(H) \times P(E|H)$.[90]

Hence,

$$P(H|E) = P(H) \times P(E|H) \div P(E).$$

Under this framework, H represents the decision-maker's hypothesis, while E stands for her evidence. The theorem integrates the probability of H prior to the arrival of E ($P(H)$), the general probability of E's presence in the world ($P(E)$), and E's probability of being present in cases featuring H as well ($P(E|H)$). These three factors allow the decision-maker to compute the posterior probability of her hypothesis: the probability of hypothesis H given evidence E.

The product rule and the complementation rule for mutually exclusive events form the mathematical foundations of the entire probability system. Failure to comply with these rules will produce serious distortions in the decision-maker's probabilistic assessment. These distortions will include over- and under-valuations of the relevant chances and prospects. Over time, these chances and prospects will materialize into actual events. The decision-maker's erroneous perception of those chances and prospects consequently will engender misguided decisions and actions.

An epistemic contest is an altogether different mode of factfinding that has its own criteria for evaluating competing factual claims. These criteria are qualitative rather than quantitative. They include the claim's internal coherence as

---

89.   *See* Thomas Bayes, *An Essay Towards Solving a Problem in the Doctrine of Chances*, *in* PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON 4–16 (1763), *available at* http://www.stat.ucla.edu/history/essay.pdf. For a modern statement of the theorem, see COHEN, *supra* note 53, at 68.

90.   Because of this inversion, some call the Bayes' Theorem the "Inversion Theorem." *See, e.g.*, WILLIAM KNEALE, PROBABILITY AND INDUCTION 129 (1949).

a sequential story with specified causes and effects. These criteria also require evidentiary support for every element of the claim. This requirement is twofold: It is not enough for the available evidence simply to verify the underlying factual claim; it must also show that rival factual scenarios are improbable. After being discredited by the evidence, these scenarios will be eliminated and will play no role in the court's final decision.

The end product of this procedure is the survival of the epistemically fittest factual account. Based on the natural reasoning process that includes coherence, causal specificity, evidential support, and other criteria articulated earlier in this Article,[91] the court will adopt the factual account that makes the most sense as a description of the relevant event. This winning account is called the "inference to the best explanation."[92] Factfinders need not apply any algorithms or formal logic to determine what this account is. The natural reasoning process will suffice.[93]

The epistemic mode of factfinding focuses on individual occurrences that are of consequence to the court's decision under applicable law. Courts using this mode try to ascertain the actual facts of the case, rather than figure out which gamble on the truth yields the highest rate of correct decisions over time. We label this mode "epistemic contest" because of its elimination procedure, whereby a claim with stronger epistemic credentials—one that scores most on coherence, causal specification, evidential support, and other criteria associated with natural reasoning—prevails over rival allegations and removes them from the factfinders' agenda. This decision rule is the key component of the epistemic mode of factfinding. It separates the epistemic mode from the gambling system under which high- and low-probability scenarios differ from each other quantitatively, but not as a matter of substance.[94]

To see this point more vividly, consider the tension between adjudicative factfinding and one of the pillars of the mathematical probability system: the complementation rule. This tension was first articulated in the famous Gatecrasher Paradox featuring 1,000 rodeo spectators, of whom only 499 paid for their admission, and no other evidence.[95] Under this somewhat artificial but illustrative setup, a preponderant 0.501 probability supports the rodeo organizers' allegation that *S*, a randomly picked spectator, is one of the gatecrashers. On the other hand, *S*'s claim that he actually paid for his admission to the rodeo only has a 0.499

---

91. *See supra* notes 61–62 and accompanying text.

92. *See* PETER LIPTON, INFERENCE TO THE BEST EXPLANATION 55–57 (2d ed., 2004); Pardo & Allen, *supra* note 55, at 227–42. For a foundational philosophical work, see Gilbert H. Harman, *The Inference to the Best Explanation*, 74 PHIL. REV. 88, 88–89 (1965).

93. *See* Allen & Leiter, *supra* note 68, at 1532–34.

94. Another big difference between the two modes of reasoning is the "indifference principle" upon which mathematical probability calculus proceeds. The epistemic mode of reasoning rejects this principle. For details, see Alex Stein, *The Flawed Probabilistic Foundation of Law and Economics*, 105 NW. U. L. REV. 199, 219–22, 236–42 (2011).

95. *See* COHEN, *supra* note 31, at 75.

probability. Hence, under the preponderance standard that generally applies in civil litigation, the organizers appear to be entitled to recover the admission money from S, which is patently absurd.[96]

The difficulty here is the small—indeed, barely visible—margin of advantage that allows the organizers to surpass the preponderance threshold, mathematically defined. Allowing the organizers to override S's defense with a 0.501 probability makes no sense at all.

Under the epistemic mode of factfinding, this difficulty does not exist. This mode of factfinding identifies the best explanation that outscores its rivals by a wide margin.[97] As we already explained, a factual scenario that this reasoning system identifies as a winner will always stand out as epistemically superior to other available scenarios. This scenario will always be more comprehensive, more coherent, and better evidenced than its rivals. In the Gatecrasher case, the rival hypotheses will be a statistic from the plaintiff and the fully fleshed out testimony of the defendant describing a perfectly plausible scenario from a cognitively competent individual with first-hand knowledge.[98]

Consider now the conjunction paradox. As we explained in the Introduction, this paradox is a logical consequence of the product rule. Under this rule, when two probabilities that equal less than one are multiplied by each other, the resulting number gets smaller than either of the two probabilities. This rule has a simple explanation: A compound two-event gamble is riskier than a gamble on one of the two events.[99] For example, when a person tosses a fair coin once, his probability of getting heads equals 0.5. When he tosses it twice, his probability of getting two heads in a row goes down to 0.25.

The relative plausibility system, however, is not a system of gambling. Under this system, "inference to the best explanation" prevails over rival inferences and removes them from the factfinder's consideration. Hence, when two distinct inferences to the best explanation support the plaintiff's allegations about the defendant's wrongdoing (W) and the harm she sustained therefrom (H), the defendant's statistical chances of not committing the alleged wrongdoing or, alternatively, not causing the alleged harm—represented by the disjunctive formula $[1 - P(W)] + [1 - P(H)] - \{[1 - P(W)] \times [1 - P(H)]\}$—become inconsequential. They are not and, indeed, cannot be present in any specific occurrence in the empirical world. Rather, they spread themselves across different occurrences that are mutually incompatible with each other. This predicament makes statistical chances epistemically inferior to case-specific inferences to the

---

96.   *Id.*

97.   When it is not wide, the burden of persuasion acts as a tiebreaker and assigns the victory to the status quo.

98.   This decisive explanation of the Gatecrasher hypothetical is novel. We believe it avoided the attention of evidence scholars in part because the relative plausibility theory arose prior to seeing that its philosophical base was inference to the best explanation. Other matters were under consideration at its genesis.

99.   *See* Stein, *supra* note 94, at 209–10 (explaining the product rule).

best explanation.[100] When these inferences take the plaintiff's case above the "preponderance" threshold, a sheer combination of chances that are not empirically present in the individual occurrence cannot undo this advantage.[101]

2. Naturalism

The relative plausibility account has the marked advantage of embracing natural reasoning processes, rather than imposing an odd epistemology on jurors and judges. Indeed, this advantage, as we saw, was a central animating feature of the Supreme Court's opinion in *Old Chief*.[102] Natural common-sense reasoning is obviously not infallible, but at the same time it has a solid epistemological base. This base incorporates ordinary people's ever-increasing capacity to tame the unruly complexity of the universe, and by doing so contribute to the survival of the species.[103] Unless there is some reason to think that this reasoning is perverse and in need of correction—and that a proposal to fix it would actually improve adjudicative factfinding—the law should align with common sense.[104] That is, the law should follow practices that have, in Alvin Goldman's words, "a comparatively favorable impact on knowledge as contrasted with error and ignorance[.]"[105]

Although probabilistic reasoning is not alien to the human mind, its relative frequency and subjective belief versions play little role in everyday reasoning about human affairs. Rather, people reason about their affairs predominantly through stories, scripts, and narrative events.[106] Jurors (and judges) are no different. They construct narratives out of the evidence and choose the best narrative as representing the best attainable approximation of the truth.[107] They do not consider the probability of various elements of a story being true, but look instead to it holistically. Evidence that confirms and disconfirms parts of the relevant story is always integrated into the story's acceptance or rejection as a whole. Moreover, what a rational person would expect to see in a story about civil or criminal liability may differ markedly from the implications of the liability's formal elements. Again, this was a critical part of the Supreme Court's decision in *Old Chief*.

To the extent one wants to advance factual accuracy at trial, it is sensible to model factfinding on the methods that proved helpful in the decision-makers' lives. There are cases in which mathematical probability takes over, but those

---

100.   *See generally* Allen, *Factual Ambiguity, supra* note 16, at 605–09.

101.   *Id. See also* Michael S. Pardo, *The Nature and Purpose of Evidence Theory*, 66 VAND. L. REV. 547, 600–10 (2013) (explaining how inference to the best explanation determines relevancy and probative value of evidence).

102.   Old Chief v. United States, 519 U.S. 172, 186–89 (2007).

103.   *See generally* Allen & Leiter, *supra* note 68.

104.   *See generally* Alex Stein, *Book Review: Are People Probabilistically Challenged?*, 111 MICH. L. REV. 855 (2013) (vindicating an ordinary person's common sense against probabilistic irrationality accusations leveled by behavioral economists).

105.   ALVIN I. GOLDMAN, KNOWLEDGE IN A SOCIAL WORLD 5 (1999).

106.   *See* Diamond et al., *supra* note 63.

107.   *See id.*

cases are exceptional. They involve rules, entitlements, and remedies that depend upon mathematical probability by design. The prime examples of such rules, entitlements, and remedies are market-share liability for defective products, doctors' liability for patients' lost chances to recover from illness, employers' liability for discriminating against classes of employees, trademark infringers' liability for consumer confusion, and the election law protection against redistricting manipulations.[108] For factual determinations in other types of cases, mathematical probability is simply irrelevant, although it may play a role as part of an expert witness's testimony that factfinders merge with the specifics of the case, as they often do with DNA evidence.[109]

Critically, natural reasoning equips decision-makers with experience-based tools that allow them to engage in a global (or holistic) assessment of evidence. These tools are necessary for resolving complexities that adjudicative factfinding routinely presents. Mathematical probability is lacking in these tools. Using mathematical probability instead of natural reasoning would therefore make factfinding hopelessly unmanageable. Consider a paradigmatic trial scenario that one of us analyzed in previous work:

> Suppose a witness begins testifying, and thus a fact finder must decide what to make of the testimony. What are some of the relevant variables? First, there are all the normal credibility issues, but consider how complicated they are. Demeanor is not just demeanor; it is instead a complex set of variables. Is the witness sweating or twitching, and if so is it through innocent nerves, the pressure of prevarication, a medical problem, or simply a distasteful habit picked up during a regrettable childhood? Does body language suggest truthfulness or evasion; is slouching evidence of lying or comfort in telling a straightforward story? Does the witness look the examiner straight in the eye, and if so is it evidence of commendable character or the confidence of an accomplished snake oil salesman? Does the voice inflection suggest the rectitude of the righteous or is it strained, and does a strained voice indicate fabrication or concern over the outcome of the case? And so on.[110]

---

108. *See* Thornburg v. Gingles, 478 U.S. 30, 52–61 (1986) (approving use of statistics for determining racially polarized voting and minority vote dilution); Schechner v. KPIX–TV, 686 F.3d 1018, 1022–25 (9th Cir. 2012) (using statistics to determine age and gender discrimination in employment); J. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION §§ 23:1–18 (4th ed. 2008) (attesting that courts rely on consumer survey statistics to determine likelihood of consumer confusion in trademark infringement suits); ARIEL PORAT & ALEX STEIN, TORT LIABILITY UNDER UNCERTAINTY 61–67, 116–29 (2001) (analyzing court decisions that used mathematical probability to determine manufacturers' market-share liability and doctors' liability for patients' lost chances to recover).

109. *Cf.* Andrea Roth, *Safety in Numbers? Deciding When DNA Alone is Enough to Convict*, 85 N.Y.U. L. REV. 1130 (2010) (explaining how DNA evidence integrates with other evidence presented in criminal trials and when it warrants a finding "beyond a reasonable doubt" upon which jurors should convict the defendant).

110. Allen, *Factual Ambiguity*, *supra* note 16, at 625–26.

In this case, and in many others, reliance on mathematical probability will create a decisional impasse or, worse, will take factfinders astray. Complexity and multiplicity of the relevant variables make them unsusceptible to accurate representation that uses mathematical language. An elegant, but inaccurate, representation of those variables will move factfinders away from the actual event and create distortions.[111]

To sum up, there is a strong fit between naturalism and adjudicative factfinding. A legal system would take upon itself a considerable risk by adopting epistemic norms that run contrary to people's normal cognitive practices. As a baseline rule, it would seem that the system ought to identify and adopt those epistemic norms that work in practice and that factfinders can use expediently to achieve the desired result: accuracy of decisions. Common sense thus remains the most attractive epistemic norm to adopt. As a renowned philosopher once put it, "[w]e need only a reasonable layman, not a logician or statistician, to determine what is beyond reasonable doubt."[112]

3. Empirical Truth

Under the relative-plausibility framework, evidence upon which factfinders identify the winning story tracks the empirical truth about the specific event in a way that frequentist probability does not. To see why, consider the following question: Would evidence that identifies the winning story have unfolded the way it did if that story were false?

Falsity of the winning story is always a possibility, as there are no facts about which factfinders can ever be certain. Yet, evidence that allows the winning story to win the plausibility contest does not come into existence by accident. This evidence must satisfy a demanding set of epistemic criteria that comprise natural reasoning about events. 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.

None of this is true about frequentist probability, which attaches indiscriminately to all factual occurrences that fall into the relevant category of events. In the much-discussed Blue Bus case,[113] for example, the fact that the Blue

---

111.   *See id.* at 626. For a somewhat similar conception of legal evidence, see DOUGLAS WALTON, LEGAL ARGUMENTATION AND EVIDENCE 200 (2002); *see also* Ronald J. Allen, *Artificial Intelligence and the Evidentiary Process: The Challenges of Formalism and Computation*, 9 ARTIFICIAL INTELLIGENCE & L. 99 (2001).

112.   L. Jonathan Cohen, *Freedom of Proof*, *in* FACTS IN LAW, 16 ARCHIVES FOR PHILOSOPHY OF LAW AND SOCIAL PHILOSOPHY 1, 21 (William Twining ed., 1983).

113.   This case is an adaptation from Smith v. Rapid Transit, Inc., 58 N.E.2d 754, 754–55 (Mass. 1945) (holding that the plaintiff had failed to make a prima facie case as to the ownership of the bus that injured her by showing that the defendant operated the only bus franchise on the street in question; explaining that "that it is 'not enough that mathematically the chances somewhat favor a proposition to be proved; for example, the fact that colored automobiles made in the current year outnumber black ones would not warrant a finding that an undescribed automobile of the current year is colored and not

Bus Company owns eighty percent of the buses in town generates a 0.8 probability for any suit filed by a person who was hit by an unidentified bus. This frequentist probability does not track the truth of any story featuring a real victim who was hit by a blue bus. Instead, it stays invariant across all such stories—those that are true, and those that are utterly false[114]—and hence it is not *sensitive* to the truth.[115] Case-specific evidence, on the other hand, is sensitive to the truth of the factual narrative it supports. Falsity of that narrative normally brings along changes in the evidence.[116]

To further illustrate this pivotal point, consider a personal injury suit supported by three witnesses. The first witness is a passer-by who testifies that he saw the defendant's car running the red light and colliding with the plaintiff's car. The second witness is a doctor who testifies about the injuries that the plaintiff sustained from the accident. The doctor tells the court that she first saw those injuries when the plaintiff came to the hospital the day after the accident. The third witness is the plaintiff himself. The plaintiff testifies that he was injured from the collision with the defendant's car but cannot tell how it happened because the accident was so sudden.

The defendant disagrees with all three witnesses. He testifies that the plaintiff's car ran the red light and collided with his vehicle. The defendant also tells the court that he saw no injuries on the plaintiff when the plaintiff came out of his car after the accident to observe the damage to the car.

In this case, the plaintiff has a coherent and causally articulated story supported by two independent and unbiased witnesses, while the defendant's story is much weaker epistemically. The plaintiff's story consequently wins the plausibility contest. Under the relative plausibility framework, the court would consequently conclude that the plaintiff proved his case by a preponderance of the evidence.

From a frequentist probability perspective, however, things are markedly different. If the probabilities of truthfulness attaching to the plaintiff's witnesses

---

black'"; and concluding that "[t]he most that can be said of the evidence in the instant case is that perhaps the mathematical chances somewhat favor the proposition that a bus of the defendant caused the accident. This was not enough." (citing Sargent v. Massachusetts Acc. Co., 29 N.E.2d 825, 827 (Mass. 1940))).

114. For that reason, the Massachusetts Supreme Judicial Court denied the plaintiff recovery in a case featuring an analogous set of facts. *See* Smith v. Rapid Transit, Inc., 58 N.E.2d 754, 755 (Mass. 1945).

115. *See* TIMOTHY WILLIAMSON, KNOWLEDGE AND ITS LIMITS 147–63 (2000); *see also* David Enoch et al., *Statistical Evidence, Sensitivity, and the Legal Value of Knowledge*, 40 PHIL. & PUB. AFF. 197, 202–10 (2012) (unfolding an interesting application of "sensitivity" to evidence law); Judith Jarvis Thomson, *Liability and Individualized Evidence*, *in* RIGHTS, RESTITUTION, AND RISK: ESSAYS IN MORAL THEORY 225 (William Parent ed., 1986) (rejecting statistical proof of adjudicative facts and unfolding the "guarantee" requirement, analogous to "sensitivity").

116. Based on a similar analysis, one of us has argued that verdicts grounded upon naked statistical evidence violate the "principle of maximal individualization"—a person's fundamental protection against risk of error. STEIN, *supra* note 1, at 91–106.

equal 0.8, 0.8, and 0.7, the probability that all of these witnesses gave truthful testimony would equal 0.45. Hence, one of the plaintiff's witnesses testifies untruthfully in 55 cases out of 100. This statistical proposition has a serious shortcoming: It holds invariant across all cases, including those in which the three witnesses tell the court nothing but the truth. This invariance makes the frequentist proposition insensitive to the truth.[117]

The plaintiff's epistemic advantage, on the other hand, is not invariant. Falsity of one of his witnesses' testimony might—and often will—alter the evidence that otherwise would allow him to prevail in the plausibility contest. The plaintiff will consequently lose the contest instead of winning it. Whether it will happen in many cases or in just a few depends on the epistemic gap separating the plaintiff's story from the defendant's account of the event. When the gap is substantial, there is every reason to believe that all of the plaintiff's witnesses are telling the truth. Factors that determine how substantial this gap is—coherence, causal articulation, consilience, evidential coverage, and others—consequently trump frequentist probability in any individual case. As the famous saying goes, "for individuals there are no statistics, and for statistics there are no individuals."[118]

The relative plausibility system gives a citizen clear signposts about her legal rights and obligations, which facilitates her compliance with the law and bargaining in the shadow of the law.[119] The system also vests the decision in a commonsensical framework of the decision-making process that parties and factfinders can easily understand and operationalize. These characteristics should lead the economically minded to predict that the conventional organization of our trial system has a positive effect on adjudication and on primary activities. Far from suggesting that this effect is socially optimal (in some kind of idealized way or in an imaginary world free from the actual constraints of reality), we suggest instead that the present operation of the proof rules is probably the best that one can do. We now move to examine this assessment by juxtaposing the relative plausibility account of juridical proof against two recent efforts to provide alternative explanations and prescriptions.

II. EVIDENCE THRESHOLDS

In this Part of the Article, we make three points. First, we describe Kaplow's model and uncover its severe practical and conceptual limitations. Second, we show an internal inconsistency in the model that eviscerates it. Third, we demonstrate that Kaplow failed to account for the interactions between burdens of proof and substantive law. This failure led Kaplow to design a complex and

---

117.   *See* Enoch et al., *supra* note 115, at 202–10.

118.   *See* George O'Brien, *Economic Relativity*, 17 J. STAT. & SOC. INQUIRY SOC'Y IR. 1, 11 (1942).

119.   *Cf.* Robert H. Mnookint & Lewis Kornhauser, *Bargaining in the Shadow of the Law: The Case of Divorce*, 88 YALE L.J. 950 (1979) (coining the phrase "bargaining in the shadow of the law").

highly unconventional factfinding apparatus that tries to attain the same objectives that the extant law already attains.

A. Do Evidence Thresholds Work?

Kaplow advises policymakers to make a sustained effort to suppress activities associated with undesirable concentrations of harm versus benefit, while allowing all other activities to take place. To this end, he argues, policymakers should formulate evidence thresholds associated with both types of activities. Courts should use those thresholds as a basis for holding defendants liable and for vindicating their behavior. Evidence thresholds should be doing what the burdens of proof presently do: They should completely replace the conventional formulation and allocation of proof burdens.

The social loss that Kaplow's model aims to minimize has two distinct components. The first component is harm caused by activities that the legal regime does not manage to suppress. The second component is the unrealized benefits from activities that the regime suppresses. To achieve optimal deterrence, the legal regime ought to minimize harm and avoid suppressing benefits as a total sum (subject to costs of enforcement). Adjudicative factfinding plays an important role in that endeavor. The burden of proof doctrine, in turn, plays a pivotal role in adjudicative factfinding. Thus, it is critical to set the burden of proof correctly.

Kaplow argues that the doctrine in its present form is completely disconnected from the pursuit of social welfare. According to Kaplow, the doctrine is so utterly out of touch that it could be replaced with a rule that authorizes courts to decide cases by tossing a coin: the ensuing cost to society would be zero.[120] This sharp criticism also claims that evidence scholars, notwithstanding their focus on burdens of proof, failed to notice this profound dysfunctionality.[121]

We disagree with Kaplow's criticisms of both the doctrine and evidence scholars and thus do not subscribe to his grim conclusion. As we will show, evidence scholars have long shared Kaplow's concern but have overlooked no dysfunctionality, for the most part because none exists. As we showed in Part I, the burden of proof doctrine does promote efficiency and fairness, and the doctrine's overall effect on society's welfare is most likely positive.

Notwithstanding the conceptual brilliance of Kaplow's proposal and the refinement and rigor of the underlying economic analysis, our overall assessment of this mechanism is unfavorable on its own terms. In what follows, we show that Kaplow overstated his mechanism's advantages and paid no attention to its disadvantages, particularly its distortionary effects and high informational costs. Consequently, replacing the conventional burden of proof doctrine with this mechanism would be inadvisable. In fact, this mechanism would be impossible to implement because it requires information that can only be obtained at a prohibitive cost. Moreover, because this information will constantly require updating, the ensuing administrative costs would be virtually infinite and hence

---

120.   *See* Kaplow, *supra* note 18, at 742–43, 749.
121.   *Id.* at 742–44.

unbearable. This reason alone calls for a rejection of Kaplow's model on strictly economic grounds.

Kaplow's evidence thresholds capture activities probabilistically associated with different concentrations of harm and benefit.[122] Each of those concentrations forms a bundle. Society must either accept or reject a concentration in its entirety by permitting or penalizing the underlying activity. By penalizing the activity, society will endeavor to suppress all of its effects, harmful and beneficial. Conversely, society's decision not to penalize the activity will allow both harmful and beneficial effects to materialize. Evidence identifying these concentrations and associated activities can thus be employed as a powerful policy instrument. Policymakers can use it to set up rules of decision specifying the permitted and prohibited concentrations of harm and benefit (the "thresholds"). These rules will instruct courts to impose liability on defendants whose conduct is probabilistically associated with a prohibited concentration of harm versus benefit. Conversely, the rules will instruct courts to exonerate defendants whose conduct falls into a permitted zone (demarcated in probabilistic terms as well). Crucially, the probability by which a defendant's conduct will be tied to a prohibited concentration of harm and benefit will be a function of the concentration's discrete mix of harms and benefits. For example, a defendant's very low probability to have committed a particular act will justify liability if, in the set of such acts, social harm greatly exceeds social benefit. The reverse might be true as well. In order to impose liability in connection with some other category of primary behavior, efficiency may require a very high probability of the act having been committed. In sum, there will be a direct relationship between the probability needed for liability and the relative dominance of benefit over harm, or vice versa, in the relevant category of primary behavior. As the behavior's probability of being socially useful on the aggregate increases, the evidence threshold for liability increases as well. Conversely, when the behavior is associated with a socially negative benefit versus harm tradeoff, the evidence threshold for liability decreases. According to Kaplow, the burden of proof should thus be set discretely for each category of primary behavior to ensure socially optimal outcomes.

Kaplow argues that courts will improve society's welfare by applying these rules.[123] One of the anticipated advantages of Kaplow's approach is adjustability. Under Kaplow's system, policymakers and courts would be able to update the evidence thresholds as they receive new information about different mixes of harms and benefits accompanying people's behavior. Kaplow acknowledges that the implementation of this system would require considerable expenditures in order to generate the information policymakers and courts will need.[124] He predicts, however, that, under certain plausible conditions, the system's benefits will outweigh the cost.[125]

---

122.   *Id.* at 781–86.
123.   *Id.*
124.   *Id.* at 786–89.
125.   *See id.*

This prediction is unwarranted. First, policymakers and courts would find it impossible to generate the information that would identify the appropriate evidence thresholds. Setting up evidence thresholds that capture all relevant categories of conduct and the accompanying concentrations of harm versus benefit is a monumental task in and of itself. Second, Kaplow's model omits a critical element: the dynamic nature of our society. Once any threshold is set, people will react to it and modify their primary behavior to exploit it as best they can. For Kaplow's model to function, it must have information about the thresholds that courts could depend upon. Critically, the model must also predict the future, which would be, in a word, difficult. We discuss these two points in turn.[126]

Kaplow's approach requires that the relevant categories of activities suitable for regulation be identified in advance; otherwise, every single act of a person would receive a unique analysis under his proposal. We need only note that this would pose insurmountable informational costs. However, any reasonable alternative will be as costly. Any "category" of primary behavior will have within it numerous subsets of activity featuring different levels of harm and benefit. Take trespass on land. One cannot set a Kaplow threshold for that category because it is too abstract. Obviously, someone walking across someone else's land without permission creates different inefficiencies (or efficiencies![127]) than a trespasser who uses someone else's land as a waste dump.[128] Nor would we want to lump the young lovers parking in a secluded, private property that belongs to another person with trespassers who dump waste. Even the cost-benefit analysis of a person taking a stroll on someone else's land differs from that of the young couple having a tryst. The category of "using someone's land as a waste dump" would not be good enough either, for just as obviously there is a big difference between dumping clean dirt and dumping toxic waste.

To operationalize Kaplow's proposal requires policymakers to articulate all these categories, along with all other forms of human activity, and gather dependable information about the mix of harms and benefits associated with each category. This is outlandish. There is no cost-effective way to collect information about the welfare implications of strolling on someone else's land, compared to having a romantic liaison while parked there. Nor is it feasible for policymakers to classify in advance the relative harms and benefits that can come from all forms of material dumped on someone else's land.[129]

---

126.   Related to our second point, evidence has distortionary effects on primary behavior, which would be exacerbated by Kaplow's evidentiary thresholds. *See* Parchomovsky & Stein, *supra* note 14.

127.   *See* Ben Depoorter, *Fair Trespass*, 111 COLUM. L. REV. 1090 (2011) (identifying circumstances in which trespass can be efficient); Gideon Parchomovsky & Alex Stein, *Reconceptualizing Trespass*, 103 Nw. U. L. Rev. 1823, 1849–58 (2009) (same).

128.   For a classic account of the economic consequences of trespass and nuisance, see Thomas W. Merrill, *Trespass, Nuisance, and the Costs of Determining Property Rights*, 14 J. LEGAL STUD. 13 (1985).

129.   Under the terminology that Kaplow developed in another article, this task would involve prohibitive promulgation costs, which makes comprehensive rule-making

Remarkably, even this insurmountable problem does not fully present the informational challenges facing the model. Kaplow's central idea is that his thresholds can be set to achieve socially optimal results. Laudable as it is, this idea neglects another theme of recent evidence and procedural scholarship that has focused on the dynamic nature of primary behavior and the complex interactions between primary and litigation behavior.[130] Although Kaplow notes the possibility of updating the thresholds, he essentially models the legal system and primary behavior as though they were static: Policymakers determine the right thresholds, courts apply them, and people will have the incentives to behave optimally.[131] Alas, "the life of the law"[132] is dynamic, not static; self-seeking actors will react to and try to exploit and avoid the thresholds in myriad ways.[133] Policymakers will have to predict these avoidance and exploitation efforts and adjust the thresholds accordingly. The amount of information that these adjustments will require is unrealistic.[134]

---

socially inferior to case-by-case adjudication. *See* Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 579–83 (1992).

130. *See* Ronald J. Allen, *Rationality and the Taming of Complexity*, 62 ALA. L. REV. 1047, 1056–59 (2011); Ronald J. Allen & Alan E. Guy, Conley *as a Special Case of Twombly and* Iqbal: *Exploring the Intersection of Evidence, Procedure, and the Nature of Rules*, 115 PENN ST. L. REV. 1, 6–7 (2010); *supra*, note 14. This literature builds on various efforts to model the legal system as a complex adaptive system. *See, e.g.*, J.B. Ruhl, *Law's Complexity: A Primer*, 24 GA. ST. U. L. REV. 885, 886 (2008); J.B. Ruhl, *The Fitness of Law: Using Complexity Theory to Describe the Evolution of Law and Society and Its Practical Meaning for Democracy*, 49 VAND. L. REV. 1407, 1410 (1996) ("[L]aw and society coexist interdependently and dynamically, approximating the behavior of nonlinear systems as they exist in the physical world."); *see also* Magda Osman, *Controlling Uncertainty: A Review of Human Behavior in Complex Dynamic Environments*, 136 PSYCHOL. BULL. 65, 65 (2010). For general overview of these efforts, see J.B. Ruhl, *Complex Adaptive Systems Literature*, SOCIETY FOR EVOLUTIONARY ANALYSIS IN LAW, https://www4.vanderbilt.edu/seal/scholarly-resources/complex-adaptive-systems-literature (last visited Feb. 11, 2013).

131. *See* Kaplow, *supra* note 18, at 752–62.

132. *See* OLIVER WENDELL HOLMES JR., THE COMMON LAW 1 (1923) ("The life of the law has not been logic: it has been experience.").

133. *See* Parchomovsky & Stein, *supra* note 14, at 526–42 (showing that rational actors align their behavior with evidentially favorable outcomes).

134. Remarkably, Kaplow claims that "the information requirements for determination of the optimal evidence threshold do not differ greatly from those for the determinants for the preponderance rule or other rules based on *ex post* likelihoods." Kaplow, *supra* note 18, at 772 n.59. His subsequent discussion tries to substantiate this surprising claim by carrying out a comparison between the information that goes into his evidence thresholds and the components of the conventional proof burdens, stated in Bayesian terms. *Id.* at 787–88. The components of the conventional proof burdens include, according to Kaplow, the frequencies with which harmful and beneficial acts are brought before courts of law and the probabilities of "evidence at threshold, given harmful/benign act." *Id.* at 781 fig.5. Needless to say, this formulation is not—and has never been—part of our law. *See supra* Part I.A. Under the conventional proof burden, a party with a story better evidenced than his opponent's story wins a civil suit; and in a criminal case, the prosecutor will secure the defendant's conviction when her story describing how the defendant

In sum, Kaplow's proposal cannot get off the ground without a mechanism that identifies not only the present mixes of harm and benefit but also how those mixes will change in response to a rule that might be imposed. We do not see what information-producing mechanism can be utilized to make Kaplow's proposal viable. Authorizing courts to make the required tradeoffs on a case-by-case basis is not viable either. The cost of making, updating, and remaking those tradeoffs is prohibitive (if not conceptually impossible). Worse yet, court-made tradeoffs will be unpredictable, which will chill beneficial conduct (along with the harmful conduct) and reduce the deterrent efficacy of the law. The twin problems of heavy information costs and unpredictability will clearly reduce the social utility of the legal system.[135]

B. Evidence Thresholds and Bayes' Theorem

We now move to discuss the integration of Kaplow's evidence thresholds into the overall probabilistic assessment of the case. Understanding this integration is important for its own sake, and it also helps us to unpack the thresholds' analytics—the idea that civil and criminal liability should be determined upon information other than the totality of the evidence.

Under Kaplow's model, the evidence thresholds will take the place of the posterior probability of the alleged violation. To the extent that courts inquire into probabilities at all, their inquiries are part of their customary pursuit of accuracy *ex post*—an approach that Kaplow categorically denounces for being unrelated to welfare.[136] According to Kaplow, relying on the posterior probabilities of

---

committed the crime is overwhelmingly better evidenced than the defendant's story. *See supra* notes 53–69 and accompanying text. Within this framework, information that factfinders need, on top of their general understanding of the world, is typically limited to the event on trial. Under Kaplow's system, on the other hand, courts will require a complete and fully updated encyclopedia of social facts. Incidentally, information that factfinders would need under the Bayesian formulation of the proof burden does not make a chapter in that encyclopedia. Under Bayes' Theorem, factfinders would only have to add to their case-specific evidence the prior distribution of rightful and false filings of civil suits and criminal indictments. Finding out what these distributions are is an onerous task, but it is far from being as onerous (and as costly) as the acquisition of data about the combinations of harms and benefits that accompany people's multifarious endeavors.

135.   Our concerns are borne out in real life, as tax law demonstrates. Tax theory suggests that the best tax is a uniform one that applies across the board. *See* Daniel Shaviro, *An Economic and Political Look at Federalism in Taxation*, 90 MICH. L. REV. 895, 964–65 (1992). Taxes attuned to the specifics of individual taxpayers' circumstances may seem attractive, but society can ill-afford the multiplicity of rules they create because it allows dishonest taxpayers to move between the rules in a quest for the most favorable tax treatment. Varying taxes are analogous to Kaplow's evidence thresholds in that they make gaming and consequent distortions of the system inevitable.

136.   *See* Kaplow, *supra* note 18, at 743–44, 747 ("It is hard to avoid the conclusion that the strong attraction of the 50% requirement is substantially attributable to its being a powerful focal point, some of its power deriving from there being no other focal points—besides 0% and 100%, neither of which has any appeal. . . . [T]here is almost no overlap between the direct determinants of the preponderance rule (or other such rules,

violations can bring about socially pernicious consequences. To illustrate, take an evidence threshold representing a socially pernicious concentration of harm and benignancy and assume that the defendant's conduct fits into this threshold. Under Kaplow's system, the court should hold the defendant liable. The traditional *ex post* approach, however, will produce a different result when additional evidence unassociated with the threshold takes the posterior probability of the alleged violation below 0.5. Under the criteria Kaplow favors, the ensuing exoneration would be good for the defendant but bad for society.

By the same token, an evidence threshold calling for the defendant's exoneration may coexist with other evidence that carries the violation's posterior probability above 0.5. Under this scenario, Kaplow's system would still favor exoneration. Specifically, it would tell us that a court decision holding the defendant liable would be bad not only for the defendant, but also for society as a whole. Such a decision would inflict on the defendant a deprivation that is socially unnecessary and individually undeserved. Society, for its part, would experience a crippling chilling effect on beneficial activities. Under this set of facts, applying the conventional approach would be good for the plaintiff but bad for society.

For this reason, Kaplow devotes considerable effort to disassociating his evidence thresholds from courts' determinations of the violation's posterior probability. He underscores that "the optimal evidence threshold could be associated with any *ex post* probability whatsoever" and that "the determinants of the [evidence threshold and the *ex post* probability] are largely unrelated, making it entirely plausible that the first could be high and the second low, the first low and the second high, and so forth."[137] This statement, together with Kaplow's illustrations of how evidence thresholds are designed to work,[138] indicates that these thresholds function similarly to likelihood ratios under Bayes' Theorem. Despite this similarity, Kaplow avoids treating the two concepts as exactly the same.[139] This noncommittal approach makes it difficult for one to understand what evidence thresholds exactly represent.

Our evaluation of Kaplow's evidence thresholds will rely on Bayes' Theorem, notwithstanding his ambivalence.[140] We do so for a number of reasons.

---

including proof beyond a reasonable doubt) and those for the optimal evidence threshold.") (footnote omitted).

    137.   *Id.* at 784–85.

    138.   *Id.* at 785–86 ("Suppose, for example, that many of the harmful type of act are committed but there are few benign acts that might be confused with it. In that case, a moderate evidence threshold would be associated with a very high likelihood that a given act before the tribunal was of the harmful type. To implement the preponderance rule may then require an extremely low evidence threshold. Now imagine instead that there are many benign acts and few harmful ones. Then, in order to raise the *ex post* likelihood from a very low level to 50%, it may be necessary to set the evidence threshold at an extremely high level.") (citation omitted).

    139.   *See id.* at 812–13.

    140.   Our understanding of the "evidence thresholds" has also been instructed by Louis Kaplow, *Likelihood Ratio Tests and Legal Decision Rules*, 15 AM. L. ECON. REV. (forthcoming 2013), *available at* http://ssrn.com/abstract=2284035.

Similarity between the thresholds and likelihood ratios is one of those reasons, but not the dominant one. Bayes' Theorem offers an analytically precise way to see how different items of evidence connect to each other as parts of an integrated appraisal of probability. The theorem also helps explain why decision-makers should base their decisions on the totality of the evidence that captures all relevant information. Finally, and perhaps most importantly for our purposes, Bayes' Theorem allows decision-makers to pinpoint the distortions that their decision would engender if they ignore some of the relevant information.

To facilitate this analysis, we will now assume that burdens of proof can be translated into conventional probabilities. We will also assume, counterfactually, that doing so accurately describes how our legal system operates. As we noted earlier in this Article, the general falsity of this assumption disposes of Kaplow's scheme as a descriptive matter.[141] Furthermore, Kaplow's scheme leads to perverse results even when one assumes that the burden of proof doctrine, as courts apply it, is modeled on mathematical probability.

Under the Bayesian framework, Kaplow's evidence thresholds are represented by the multiplier that transforms the hypothesis's prior probability into the posterior one: $P(E|H) \div P(E)$. This multiplier is called the "likelihood ratio." It measures the frequency with which E appears in cases featuring H, relative to the frequency of E's appearance in all cases.[142] Under Kaplow's system, H will represent the activity's harms and benefits that form the concentration evidenced by E. The question arising in connection with this system is why should the decision-maker not consider the prior probability of H? This probability represents the general chance of the given activity to produce the concentration in question. The previously unaccounted evidence (E) must update this probability, but the updating does not make the probability disappear into thin air.

Consider Kaplow's prime example of how his evidential thresholds are supposed to work.[143] The example features a recurrent scenario in which doctors try to diagnose patients with a particular disease.[144] The doctors run different tests to find out patients' scores, with "higher scores indicating a greater likelihood that the disease in question is present."[145] Under this diagnostic system, patients who have the disease show scores clustering toward the high end, whereas patients with no disease show low-end scores. The doctors' challenge, as Kaplow formulates it, is "to choose a cutoff or threshold, above which treatment will be applied."[146] To doctors wishing to avoid false positive diagnoses he recommends a high threshold.[147]

Assume now that doctors decided to follow Kaplow's recommendation and set a high threshold for patients' scores. Assume further that, on average, 99

141. *See supra* notes 56–57 and accompanying text.
142. *See* Stein, *supra* note 94, at 211–13.
143. Kaplow, *supra* note 18, at 756 (explaining the threshold mechanism).
144. Remarkably, Kaplow provides no illustrations featuring a legal proceeding.
145. Kaplow, *supra* note 18, at 756.
146. *Id.*
147. *Id.*

out of 100 patients who fall within this threshold indeed have the disease. What does it mean for a friend of yours who finds herself among the 100 patients who scored high? Is she nearly certain to have the disease? Not necessarily. To find out how bad your friend's situation really is you need to know the prior probability of her affliction. That is, you need to know the disease's recurrence in the population to which your friend belongs. Assume that this recurrence is 1 out of 1,000. That is, among every 1,000 people not yet diagnosed, only one person actually has the disease. Testing these 1,000 people under Kaplow's evidence threshold will identify that person, as he will certainly score high. The rate of false positives, however, would still be 1%, which means that 10 healthy people out of 1,000 will be falsely diagnosed with having the disease. These healthy people will be in the same pool with the person who has the disease. Your friend's probability of actually having the disease will consequently equal 1/11 (0.09).[148] With this probability, she will have a lot to worry about, but she would not need to be as anxious as a person whose probability of having a serious disease is 99/100. Hence, prior probability (represented in this example by the disease's recurrence in the relevant population) can move the case from one concentration of harm versus benefit to an altogether different concentration. From a social welfare perspective, one of those concentrations may be acceptable and another unacceptable.

This example shows why any significant probabilistic decision should be based upon posterior probability that accounts for the totality of relevant information. Evidence thresholds alone will not do. By the same token, assignment of legal liability ought to proceed upon all relevant evidence. This evidence should include prior probabilities that attach to the socially favored and disfavored concentrations of harm versus benefit. Failure to take those probabilities into account would keep policymakers away from the posterior probability they need. Policymakers would consequently have a distorted picture of the relevant welfare implications.

Kaplow's evidence thresholds thus cannot be determined by likelihood ratios alone. They must integrate relevant prior probabilities in order to avoid distortions. For example, when the prior probability of the relevant transgression is low—say, 0.1—and the desired probability threshold for imposing liability needs to be above preponderance (>0.5), the transgression evidence (that updates the prior probability) must be exceedingly strong. The likelihood ratio that it must generate must be greater than five. The probability of having this evidence in the event of a transgression consequently needs to be more than five times greater than the probability of the evidence's general presence in the world.

Combining evidence thresholds with the relevant prior probabilities will yield the posterior probabilities upon which courts should base their assignments

---

148.   SIMON BLACKBURN, THINK 218–19 (1999). Under Bayes' Theorem, this probability is calculated by multiplying the prior odds (1/1,000) by the likelihood ratio (99/1). The posterior odds consequently become 99:1,000, which means that, of any 1,099 individuals, 99 people will and 1,000 people will not have the disease. With all other things being equal, your friend's posterior probability of having the disease will thus amount to 99/1,099, i.e., 0.09.

of legal liability. Hence, *probability* thresholds for courts' decisions are crucial despite Kaplow's attempt at undercutting their significance, and this is precisely how burdens of persuasion presently operate (without using mathematical language). The extant burden of proof doctrine utilizes posterior probabilities in the various decision rules that it requires courts to follow. Specifically, the doctrine requires courts to decide cases by juxtaposing the posterior probability of the alleged transgression against the applicable probability threshold: "preponderance," "clear and convincing," or "beyond a reasonable doubt."[149]

Kaplow has only one plausible response to this critique. This response would acknowledge the need to make liability decisions based on the totality of evidence and agree to add prior probabilities to the evidence thresholds—an addition that would produce posterior probabilities. The need to make this alteration shows that Kaplow's model, as it currently stands, may actually misfire and do a disservice to society's welfare.

Indeed, Kaplow's theory seems to contain a contradiction. As we have shown, courts can plausibly advance social welfare only by deciding on the basis of all the evidence, which is precisely the feature of the present rules that Kaplow wants to discard. Altering Kaplow's model to accommodate the implications of this analysis would bring his system close, if not make it identical, to the current legal regime. Under this regime, substantive law determines the prohibited and permitted mixes of harm and benefit. Burdens of proof, in turn, determine the probabilities upon which courts will connect individuals' actions to those mixes. Unfortunately, Kaplow's theory pays no attention to this crucial synergy and its implications for social welfare.

We turn next to a more careful look at the synergy between substantive law and the burden of proof. Our analysis of this issue demonstrates that Kaplow's system has no advantages over extant law. Indeed, for reasons given below, the opposite is true: The current legal regime is superior to Kaplow's system in many important respects.

C. Substantive Law and the Burden of Proof

Concentrations of harm versus benefit are, in fact, accounted for by our legal system. They animate our system's substantive rules of civil and criminal liability. Kaplow criticizes the burden of proof doctrine for being unrelated to these concentrations and their implications for welfare, but this critique pays no attention to the rules of tort, contract, and criminal liability that the doctrine was set up to implement. The burden of proof doctrine operates in tandem with the substantive rules of liability: It does so by setting up the probability thresholds that courts use in ascertaining the presence of the relevant factual characteristics and circumstances of individual conduct. These characteristics and circumstances place the relevant conduct in or outside the prohibited zone or, in Kaplow's terms, the socially favored or disfavored concentration of harm versus benefit. By setting up probability thresholds, the burden of proof doctrine determines the level of

---

149.   *See supra* notes 2–7.

enforcement for liability rules and where the risk of erroneous enforcement should fall. This fundamental feature of the doctrine accounts for its categorization as "substantive law" for purposes of diversity rules and various constitutional protections.[150] Kaplow pays no attention to this synergy and evaluates the burden of proof doctrine as a freestanding set of rules. This view of the doctrine cannot be right: It distorts the understanding of our entire legal system.

Take criminal prohibitions first. Criminal prohibitions capture conduct that is obviously associated with socially undesirable concentrations of harm versus benefit. Most of those concentrations include serious harm and no benefits whatsoever: Think of murder, rape, robbery, arson, theft, burglary and other serious offenses. Other concentrations that fall under the criminal law are not as malignant as the *mala per se* crimes, but they too feature substantial amounts of harm. Tax offenses, insider trading, and license violations are good illustrations of those less severe, but still criminal, concentrations of harm versus benefit.

Torts and breaches of contract have similar structures. Conduct that our system characterizes as torts or breaches of contract yields a socially negative tradeoff of harm versus benefit. Furthermore, the general negligence doctrine that controls the majority of tort cases expressly targets conduct that produces this negative tradeoff by requiring courts to impose liability on defendants whose conduct falls into a welfare-diminishing concentration of harm versus benefit.[151] Imposition of strict liability under the "cheapest cost-avoider" criterion and other formulations follows the same logic. Similar to the negligence doctrine, the rules of strict liability promote three goals: They encourage actors to exercise cost-efficient precautions against harm, while trying to reduce the cost of litigation and avoid the chilling of beneficial activities.[152]

Importantly, liability rules that apply in tort and criminal law set up an *ex ante* standard for information upon which actors make decisions about their primary activities. Under the negligence doctrine, whether the defendant acted negligently is a function of the accident's *ex ante* probability, the harm generally associated with similar accidents, and the precautions against the harm available before the accident.[153] The only *ex post* information that courts consider is the plaintiff's individual harm, and there is a good reason for that as well.[154] The requirement that the plaintiff's damage be foreseeable further entrenches the *ex ante* standard. So do various evidentiary rules such as the exclusion of "subsequent

---

150.   *See* Stein, *supra* note 8, at 79–82.

151.   *See* RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 213–17 (8th ed. 2011).

152.   *Id.* at 226–27. For a classic account, see GUIDO CALABRESI, THE COSTS OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS 26–31, 95–129 (1970).

153.   *See* RICHARD A. EPSTEIN, TORTS § 5.16 at 129–30 (1999).

154.   Allowing recovery for individual harms motivates plaintiffs best positioned to sue the wrongdoer to file suits. *See* Stein, *Two Wrongs, supra* note 36, at 1219–20.

remedial measures" evidence[155] and, in medical malpractice cases, the "error in judgment" instruction.[156]

Criminal law adopts a similar *ex ante* approach by setting up a stringent *mens rea* requirement for convictions. Under this requirement, a person acting in a way that is legally criminal is not automatically guilty of the underlying crime. The person is only guilty of a crime when he acts while being aware of the action's circumstances and probable effects.[157] The person must either have affirmative knowledge of these effects and circumstances or at least form a suspicion about their presence. Absent such knowledge or suspicion (sometimes identified as "willful blindness"), the person would normally be considered innocent.[158]

The *mens rea* requirement allows individuals to steer away from criminal liability by relying on the information they have or can easily access. Contemplating a permitted activity on the basis of that information allows a person to stay on the right side of the line separating criminal from noncriminal behaviors. The *mens rea* requirement thus reduces the cost of information for people who try to avoid criminal liability.[159] By doing so, it removes the potential chilling effect from a multitude of activities that are socially beneficial.[160]

Contract law's *ex ante* approach to information animates a number of similar rules. The most fundamental is the "bargain principle": a set of rules that give effect and attach legal consequences to the parties' mutual undertakings.[161] This principle requires courts to interpret these undertakings by identifying the parties' intent—expectations from the agreement, formed on the basis of information available at the time. The same *ex ante* information determines whether the agreement was formed by mistake or through misrepresentation and whether an unanticipated event frustrated the agreement.

As we already indicated, the burden of proof doctrine operates in tandem with these rules.[162] The doctrine performs a twofold function: It determines the level of enforcement for liability rules and allocates the risk of error in courts'

---

155.   *See* FED. R. EVID. 407.

156.   *See, e.g.*, Smith v. Finch, 681 S.E.2d 147, 149–50 (Ga. 2009) ("[I]t is well recognized that 'an after-the-fact assessment of facts or evidence cannot be the basis of a negligence claim "so long as the initial assessment was made in accordance with the reasonable standards of medical care."'" (quoting Holbrook v. Fokes, 393 S.E.2d 718, 719 (Ga. Ct. App. 1990))).

157.   For the basic *mens rea* requirement and its economic rationale, see Alex Stein, *Corrupt Intentions: Bribery, Unlawful Gratuity, and Honest-Services Fraud*, 75 LAW & CONTEMP. PROBS. 61, 67–68 (2012).

158.   *See generally* Darryl K. Brown, *Criminal Law Reform and the Persistence of Strict Liability*, 62 DUKE L.J. 285 (2012) (analyzing various forms of *mens rea* requirements and exceptions thereto).

159.   *See* Jeffrey S. Parker, *The Economics of Mens Rea*, 79 VA. L. REV. 741, 769–77 (1993).

160.   *Id.*

161.   *See generally* Melvin Aaron Eisenberg, *The Bargain Principle and Its Limits*, 95 HARV. L. REV. 741 (1982).

162.   *See supra* Part II.C.

decisions that enforce those rules.[163] Courts' applications of the doctrine produce liability decisions that associate parties' actions with the favored and disfavored concentrations of harm versus benefit. We therefore disagree with Kaplow's description of the burden of proof doctrine as disconnected from society's welfare. The doctrine does promote welfare, albeit not alone. It does so together with the substantive rules of tort, criminal, and contractual liability. Kaplow pays no attention to this synergy. Indeed, his theory completely disregards the connection between evidentiary rules and substantive law.

The general proof requirement for civil cases—preponderance of the evidence—performs an important role in enforcing the law. Under certain conditions, this requirement allows courts to maximize the total number of correctly decided cases.[164] When that happens, the number of decisions that miscategorize harmful conduct as beneficial, and vice versa, decreases as well. Moreover, as we elaborate below, when there is reason to think that the standard rule may not produce these results, adjustments are made, such as with the res ipsa loquitur presumption.[165] Other standards of proof are not calibrated to achieve this accuracy-maximizing and welfare-improving consequence. This effect of the preponderance requirement is well recognized in the law and economics literature and has a simple formal proof.[166] Contrary to Kaplow's assessment, the preponderance requirement offers our legal system much more than a "focal point."[167] When the substantive law correctly identifies activities that should be sanctioned, the most efficacious proof rules are those that permit the most accurate ascertainment of the facts (subject to cost). These rules will effectively deter potential transgressors by making them believe that society will respond to a transgression by levying sanctions sufficient to offset their private gain.[168]

The rule requiring criminal accusations to be proven beyond a reasonable doubt protects defendants against erroneous conviction (again, under certain reasonable assumptions). The rule sets up this protection by increasing the rate of mistaken exonerations of guilty criminals. This tradeoff rests on the generally accepted (albeit debatable) premise that an erroneous conviction produces greater harm than an erroneous acquittal.[169] By reducing the incidents of erroneous

---

163.　*See* Addington v. Texas, 441 U.S. 418, 423 (1979); *In re* Winship, 397 U.S. 358, 370–71 (1970) (Harlan, J., concurring).

164.　*See* STEIN, *supra* note 1, at 143–49.

165.　*See* PORAT & STEIN, *supra* note 108, at 84–95 (explaining the res ipsa loquitur doctrine).

166.　*See* STEIN, *supra* note 1, at 143–49.

167.　Kaplow, *supra* note 18, at 743. The decision rule for balanced cases mandates dismissal of the plaintiff's suit to eliminate the enforcement costs that would be expended if the plaintiff were to be awarded recovery. *See* STEIN, *supra* note 1, at 145. This rule also discourages filing of insufficiently evidenced suits. *See* Ralph K. Winter, Jr., *The Jury and the Risk of Non-Persuasion*, 5 LAW & SOC'Y REV. 335, 337 (1971).

168.　Another part of the evidence literature has identified additional complexities in the relationship between litigation and primary behavior. *See, e.g.*, Allen, *supra* note 130, at 1056.

169.　*See* STEIN, *supra* note 1, at 148–49.

convictions, the "beyond a reasonable doubt" requirement makes an additional contribution to social welfare. Erroneous impositions of criminal liability do not merely harm innocents and chill socially beneficial conducts, they also erode the difference between complying and not complying with the law. This erosion dilutes individuals' incentive to prefer a noncriminal activity over a criminal one.[170]

Kaplow argues, counter-intuitively, that the "beyond a reasonable doubt" requirement (and any other stringent evidentiary requirement for convictions) might, in fact, increase the rate of erroneous convictions.[171] Echoing a point discussed in evidence literature, Kaplow estimates that the reduced chilling of criminal-looking but benign acts and the ensuing shift in the flow of cases into the court system will result in more innocent defendants standing trial.[172] Hence, "[h]olding constant the rates of finding liability for each type of act . . . this phenomenon would increase the likelihood that individuals found liable would be ones who committed benign acts rather than harmful ones."[173]

This claim is not only counter-intuitive, but it is also at odds with the "rational actor" assumption that undergirds law and economics. There is no reason to think that the rates of finding liability for each type of act would remain constant when the prior probability of guilt for indicted defendants varies. The increase in the number of cases in which innocent defendants face criminal accusations reduces the prior probability of guilt, making it more difficult to convict, and thus bringing the protection against wrongful convictions back to its normal level.

Kaplow's theory utilizes the burden of proof mechanism to set up optimal incentives for activities generalized as harmful, on the one hand, and benign or beneficial on the other hand.[174] This generalized view of people's conduct omits from consideration the contexts within which that conduct takes place. Of these contexts, the most basic are contracts, crimes, and accidents. Failure to situate the relevant conduct—harmful or beneficial—within context creates a picture of the legal system in which too many important details are missing.

Consider contracts first. Contracts generally, and business agreements in particular, anticipate the prospect of litigation. The intensity of this anticipation determines whether the agreement will include a provision specifying the evidence that will determine performance and breach.[175] Strong anticipation of litigation will drive the parties to incorporate such a provision in their agreement. Otherwise, the parties will stay with the default rule set by the law. This rule requires the performing party to prove the alleged breach of the agreement, while shifting to

---

170.   *See* A. Mitchell Polinsky & Steven Shavell, *The Economic Theory of Public Enforcement of Law*, 38 J. ECON. LITERATURE 45, 60–62 (2000).
171.   Kaplow, *supra* note 18, at 790–91.
172.   *Id. See also* Allen, *supra* note 130, at 1056–60.
173.   Kaplow, *supra* note 18, at 791.
174.   *See id.* at 741 ("This Article explores how to set the evidence threshold in the manner that best advances social welfare.") (footnote omitted).
175.   *See* Scott & Triantis, *supra* note 43.

the nonperformer the burden to prove an affirmative defense: mistake, misrepresentation, frustration, unconscionability, and so forth. For both parties, the burden requires them to establish the relevant allegations by a preponderance of the evidence.

This legal framework makes Kaplow's evidence thresholds completely irrelevant. In the contract context, all the burden of proof doctrine needs to do is set up a default that best promotes the parties' exchange. For transactions that anticipate litigation, this default can be anything whatsoever, as the parties will negotiate and agree upon the desired evidentiary mechanism.[176] The default will not affect the cost of these negotiations because they will take place anyway. For other transactions, setting up evidence thresholds is not an option. Parties not anticipating litigation are generally unable to preapprove any specific document or other evidence that could decisively resolve the issues of breach and performance. Under such circumstances, one can hardly think of a policymaker, court, or other expert who could perform the preapproval function for the parties. If so, a requirement that allows a party to prevail by establishing her case by a preponderance of the evidence makes perfect sense. This requirement approximates the parties' agreement on the burden of proof had they been required to make such an agreement expressly. Furthermore, as we already mentioned, this requirement is best suited to maximize the accuracy of court decisions.[177]

In the context of criminal law, Kaplow's proposal is particularly difficult to implement. Criminal law, as we already mentioned, proceeds on the assumption that conviction and punishment of an innocent person bring about considerable harm to the person and to society in general. Arguably, this harm is much greater than society's harm from wrongful acquittals. The valuation of the harm requires policymakers to set up a very high probability threshold for criminal convictions. As a result, many guilty criminals escape prosecution and conviction. As Gary Becker first noted, to fix the resulting shortfall in deterrence, policymakers increase punishments for crimes.[178]

From an optimal deterrence perspective, Becker's model is superior to any system of law enforcement that requires police, prosecutors, and courts to expend extensive resources. Analogously, it is also superior to Kaplow's system, whose proper functioning depends on the policymakers' costly efforts to obtain information about different concentrations of harm and benefit, and associated conduct. Furthermore, introduction of Kaplow's evidence thresholds into criminal law would require our legal system to abandon its traditional protection of innocent defendants against erroneous conviction. Kaplow believes that this special protection has no place in the normatively correct tradeoff of utilities.[179] Even if he is right, which we doubt, when the legal system can adequately fend off crime without remaking its conceptual and moral foundations, there is no advantage to Kaplow's overhaul proposal.

---

176.   *Id.*
177.   *See supra* notes 164–166 and accompanying text.
178.   *See* Becker, *supra* note 47, at 194.
179.   *See* Kaplow, *supra* note 18, at 744, 798, 808–12.

III. COMPARATIVE PROBABILITY

In this part of the Article, we evaluate Cheng's system of comparative probability. Cheng attempts to eradicate the difference between the epistemic and the gambling modes of reasoning by introducing a "story" requirement for defendants.[180] Cheng's system does not allow defendants to base their defense on the aggregated probability of the scenario in which *one* of the plaintiff's allegations does not hold.[181] Against these allegations, the defendant would have to pit a story that would give the factfinders his version of the relevant events. The factfinders will then determine the probabilities of the two stories and accord victory to the story with the highest probability. Under his system, when a plaintiff's story about how the defendant breached an agreement that the two previously made has a probability of, say, 0.7, the factfinders will not automatically credit the defendant with a 0.3 probability of being right. Rather, the factfinders will give the defendant's story the probability that they think it deserves. At a maximum, this probability may get to 0.3, but 0.3 is the figure that captures every possible story inconsistent with the plaintiff's account of the events, whereas the defendant is only entitled to have one story.[182] For that reason, the probability of the defendant's story may be lower than 0.3; and so the sum of the parties' probabilities will not always add up to 1. Subsequently, the factfinders will consider and assign probabilities to the plaintiff's and defendant's stories about the damage the plaintiff suffered as a consequence of the defendant's failure to perform. Once again, the highest probability will identify the winner. Importantly, here too the factfinders will not allow the defendant to aggregate the chances of one of his stories being true.[183]

In the paragraphs ahead, we show that Cheng's override principle and mathematical probability are fundamentally incompatible. Bringing them together will engender distortions in courts' decisions. We also identify a serious analytical gap in Cheng's comparative probability system.

A. Tinkering with Conjunctions

For Cheng, the conjunction paradox is the biggest problem of mathematical probability as applied to adjudication. He believes that removing it will make trial by mathematics normatively attractive (and operationally feasible as well). The conjunction paradox is straightforward.[184] Under extant law, when a plaintiff's suit consists of two mutually independent elements, and the probability of each of those elements is 0.7, the plaintiff should win the case. However, under the product rule, the aggregate probability of these elements is 0.49—just below

---

180.   *See* Cheng, *supra* note 19, at 1259–65.

181.   *Id.* at 1263–65.

182.   Under Cheng's system, when a defendant proffers several alternative stories, he will not be entitled to rely on those stories' conjunctive probability. Rather, each story will be juxtaposed against the plaintiff's story, which, of course, motivates the defendant to come up with a single account of the events that he considers most probable. *See id.*

183.   *See id.* at 1263–64.

184.   *Id.* at 1263.

the requisite preponderance threshold (0.5). Mathematical probability tells us that the plaintiff's evidence failed to satisfy the preponderance requirement, so the plaintiff should lose the case if the legal system takes the preponderance requirement seriously enough. Indeed, the aggregated probability of the defendant's counter-allegations is 0.51 $(0.3 + 0.3 - 0.3^2)$—above the preponderance threshold. Put differently, the probability of at least one of the elements of the plaintiff's cause of action being false is 0.51. If many such cases were decided in plaintiffs' favor, there would tend to be more incorrect than correct decisions in the ratio of 51 to 49.

Cheng criticizes this straightforward analysis. He believes that this analysis misinterprets the civil burden of proof. According to Cheng, the civil proof burden does not require a plaintiff to establish that the aggregated probability of her case is greater than 0.5. All the plaintiff needs to do is show that her allegations outscore the defendant's on a scale of probabilities' ratios. In our example, these ratios are 0.7/0.3 and 0.7/0.3. They tell us that the plaintiff wins the case. Mathematical probability wins as well: The removal of the conjunction paradox reinstates its status as arguably the most logical and rigorous way to determine uncertain facts in legal disputes. Cheng's reconceptualization of the proof burden also makes peace between the gambling and the epistemic mode of factfinding. Under his system, courts will decide cases by a one-to-one comparison between the probabilities of the parties' allegations with regard to each element of the suit.[185] This comparative system, Cheng argues, aligns with factfinding that uses inference to the best explanation.[186] Under both frameworks, courts will decide cases by the relative—not absolute—plausibility of the parties' competing allegations.

Cheng's solution of the conjunction paradox has a substantive component as well.[187] As we already mentioned, he argues that a defendant "may not simply be a contrarian."[188] Specifically, parties should not be allowed to aggregate the probabilities of mutually inconsistent allegations. A tort defendant, for example, should not be permitted to say: "Contrary to the plaintiff's allegations, I did remove the ice from my doorstep; but if I did not remove it somehow, then the plaintiff sustained no injury from falling on my ice." Making a rule that disallows such counterfactual allegations will revamp the burden of proof doctrine.[189]

This solution of the paradox is close to the argument that one of us has advanced and subsequently changed. More than a decade ago, Stein argued that the conjunction paradox is unreal because a plaintiff is entitled to a declaratory judgment in her favor on each and every element of her suit that has a probability greater than 0.5.[190] Later on, he substituted this solution with a different one.[191]

---

185.   *Id.* at 1259–64.
186.   *Id.* at 1262, 1265–66.
187.   *Id.* at 1262.
188.   *Id.*
189.   *Id.* at 1262–66.
190.   Stein, *Uncertainty and Fact-Finding, supra* note 36, at 311 n.27.

Stein's reason for making that substitution was simple, although not immediately apparent: 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. More precisely, by making a legal determination that suppresses the product rule, a policymaker or court will abolish the rule, but will not make the underlying statistical consequence disappear. Take a suit that has two elements: the defendant's running the red light with his car and the resulting accident. Assume that the court uses statistical evidence and finds out that the probability of each of these elements (that we assume to be independent of each other) is 0.7. The court decides to suppress the product rule, which allows the plaintiff to win the case. The plaintiff's victory may be well deserved. However, it cannot overturn the statistical reality: The defendant's probability of being right in *one* of his allegations is 0.51. There is no way to remove this preponderance by fiat. The court, of course, may decide to ignore it, but doing so in 100 similar cases will likely produce 51 erroneous decisions and only 49 correct decisions.

This consequence foils Cheng's system as well. Cheng's system allows an event's higher probability (say, 0.7) to drive its rival probability (0.3) into nonexistence, but this override is artificial and arbitrary. If these probabilities represent what they are supposed to represent—frequencies, without more—they stand on the same informational base. Their epistemic credentials—again, things such as coherence, causal specificity, and evidential support—are identical. The only difference between these probabilities is the relative frequency of events to which each of them attests: One of those frequencies is 0.7. The other is 0.3. However, the fact that 0.7 is greater than 0.3 does not make the lower frequency nonexistent or inconsequential. Both frequencies transform into real facts over time. Under this framework, an event's probability that equals 0.7 is a consequence of the correlative probability, 0.3, which attaches to the claim that the event did not (or will not) occur. The probability that two such events will occur together (assuming again that they are mutually independent) therefore equals 0.49. This probability is a consequence of the correlative probability, 0.51, which attaches to the scenario in which one of the two events does not materialize. The latter probability, 0.51, identifies the frequency of the underlying compound event. Ignoring this frequency will not make it disappear.

The epistemic mode of factfinding does not face this predicament. The reason is simple: When an inference to the best explanation overrides its rivals and removes them from the scene, the result is neither arbitrary nor artificial. Rather, it singles out a factual scenario that outscores all the rest on the variables that inform plausibility, such as coherence, causal specificity, and evidential support. This winning scenario stands on a qualitatively superior informational platform and has better epistemic credentials than its rivals. Scenarios that it brushes aside still have positive probabilities—chances of occurrence that materialize over time. These scenarios, however, are epistemically inferior for the case at hand, which is all the

---

191.   STEIN, *supra* note 1, at 49–56; Stein, *Two Wrongs, supra* note 36, at 1233. This solution was criticized by Stein's present co-author. *See* Allen & Jehl, *supra* note 36, at 919–29; Allen, *supra* note 36, at 225–26.

decision-maker cares about. The decision-maker consequently can write these scenarios off. To be sure, this mode of factfinding is not foolproof, but it offers the most promising way to get to the truth in an individual case.

Whether one agrees with this assessment or not, it is important to understand why reconciliation between relative plausibility and the mathematical probability approach is difficult, if not altogether impossible. This point is critical. To illustrate, revisit the personal injury suit example that we developed in Part I. In that example, the court rules for the plaintiff after finding his narrative more plausible than the defendant's. As we saw, this finding is well reasoned. The plaintiff's narrative—supported by two independent witnesses, the passer-by and the doctor—outscores the defendant's claims on coherence, causal specificity, evidential support, and so forth. This qualitative advantage makes the plaintiff's narrative an epistemically superior inference to the best explanation.

Assume now that the court decides the case by using mathematical probability. The court uses the testimony of the passer-by to determine that the defendant ran the red light. The court assigns to that testimony a 0.8 probability of being true. The court evaluates the extent of the plaintiff's injuries and corresponding compensation amount by relying on the doctor's testimony, to which it assigns a 0.8 probability of being correct. The plaintiff's testimony on the issue of causation receives from the court a 0.7 probability of being true. The aggregate probability of the plaintiff's case consequently equals 0.45. Because this probability falls below the preponderance threshold, the court dismisses the suit.

This mathematical decision would be impeccable if the plaintiff's evidence had no qualitative epistemic advantage over the defendant's evidence. For example, if the plaintiff relied solely on the frequencies of the relevant events, and those frequencies were—as the court determined them—0.8, 0.8, and 0.7, the court would then also have to attest that the probabilities supporting the defendant's allegations are 0.2, 0.2, and 0.3. Under such circumstances, the court would have to allow the defendant to benefit from the implications of these correlative probabilities. Otherwise, the number of incorrect decisions that the court will deliver over time will exceed the number of correct decisions. The aggregate probability of the plaintiff's case, 0.45, should consequently decide the case, and the court will do well to decline Cheng's invitation to suppress the product rule.

But, the plaintiff's account of the events is epistemically better than the defendant's account. The plaintiff's evidence is qualitatively superior to the defendant's testimony. The defendant's testimony does not explain away his own motivation to lie in court, nor does it negate the possibility that the defendant did not see the plaintiff's injuries (which could be invisible). The plaintiff's evidence, on the other hand, gives a fully explained, coherent, and specified account of the events, in part, because it removes a potential fabrication suspicion by relying on an objective eyewitness (the passer-by) and on a medical expert whose testimony can be verified. The plaintiff's evidence therefore leads to the obvious inference that his account of the accident and its consequences is superior to the defendant's, which is all that "inference to the best explanation" entails.

        The court's mathematical decision thus amounts to a bad mistake. The court's conversion of the plaintiff's evidence into statistical frequencies fails to account for the qualitative superiority of that evidence. Indeed, the court's mathematical decision erases this superiority. By assigning mathematical probabilities to the plaintiff's allegations concerning negligence, causation, and damage, the court automatically assigns correlative probabilities to the defendant's counter-allegations. But evidence that supports the defendant's counter-allegations—his own testimony—is not commensurate with the plaintiff's evidence as a source of information from which factfinders can derive the explanations they need. The defendant's evidence bypasses a crucial credibility issue, whereas the plaintiff's evidence covers all the bases by giving the factfinders the information they need. Mathematically minded decision-makers may decide to use this epistemic difference to further discount the probability of the defendant's allegations. This discounting may not be a bad decision, but it will not remove the conjunction paradox for all cases. An epistemically better decision is to allow the plaintiff's evidence to override the defendant's testimony completely. Unlike Cheng's theory, this override is neither arbitrary nor artificial.

        Cheng's system contains an internal analytical problem as well. Under Cheng's system, a plaintiff must show probability ratios that are better than the defendant's for each element of the suit. Hence, when the plaintiff's suit has two elements with probabilities amounting to 0.9 and 0.4, and the defendant's probabilities are their reciprocals, the defendant will win the case. The plaintiff thus gets no credit for his overwhelming advantage on the first element of the suit (0.9 against 0.1). A genuinely comparative system, however, should give this credit to the plaintiff. The plaintiff's overwhelming advantage on the first element makes the overall probability of his case (0.9 × 0.4 = 0.36) six times higher than the overall probability of the defendant's case (0.1 × 0.6 = 0.06). This anomaly will be present in various cases featuring two different margins of victory—the plaintiff's and the defendant's—on two discrete elements of the dispute, regardless of whether the probabilities of those elements add up to one.[192] Cheng's system thus recapitulates the very problem he was attempting to avoid, leaving out only that the plaintiff must meet a certain threshold of greater than 0.5 on every element of the cause of action. This bizarre consequence may not be very significant in and of itself, but it reveals the system's artificiality and arbitrariness. This system makes a purely mathematical move to avoid the conjunction paradox, and nothing more. It brings about no improvements in the accuracy of court decisions.

---

        192.   Take a factfinder that assigns a 0.8 probability to the plaintiff's story concerning one element of the suit and a 0.1 probability to the defendant's story about the same element. The factfinder also assigns a 0.4 probability to the plaintiff's story about another element, while giving the defendant's competing story a slightly higher probability: say, 0.5. The plaintiff's combined story will thus have a 0.32 probability of being true. The defendant's combined story will have a much lower probability: 0.05. Under this set of facts, the plaintiff's combined story will be six times more likely than the defendant's, but Cheng's system would nonetheless accord the defendant victory. This outcome does not respect Cheng's comparative judgment criterion.

The relative plausibility system operates seamlessly in this regard as well. By requiring each party to put forward and prove an integrated story, the system handles the logical problems that arise in factfinding by allocating them to both sides of the case.[193] Specifically, it provides that any such problem goes with the story in connection with which it arises. When the party to whom the story belongs is unable to solve the problem, the story loses points in the plausibility contest. This solution markedly differs from Cheng's because it is neither arbitrary nor artificial. Indeed, this solution aligns with epistemology and common sense.

B. Law, Science, and Probability

Cheng's article makes a robust observation about the role of mathematical probability in science and in law. This observation is animated in part by Cheng's belief that "probabilistic models of inference have been incredibly successful in science, leading to dramatic insights[.]"[194] Driven by this belief, he asks, "how could statistics, a dominant modern field addressing the issue of inference, have little to contribute to proper decisionmaking in the legal system?"[195] To Cheng, "[s]uch a state of the world seems both odd and highly improbable."[196]

Cheng's canonization of mathematical probability is flawed in two respects. First, his concept of "science" is extraordinarily narrow. Moreover, Cheng pays no attention to the peculiar nature of decision-making that takes place in our courts. This decision-making has virtually nothing in common with Cheng's narrow concept of science. We discuss the two points in turn.

Mathematics and probability theory have played a critically important role in advancing knowledge in some areas of science. This is particularly true of the "King of Sciences"—physics—and even more so of high-energy particle physics. No better demonstration of this can be made than the recent announcement that the Higgs boson—or a closely related family member—has been "found" at CERN's Large Hadron Collider.[197] This experiment provides a paradigmatic illustration of how mathematical probability can advance scientific discovery. The experiment was one, large relative-frequency study with huge amounts of data analyzed probabilistically. Based on this study, the experimenters were able to attest, with a high degree of confidence, that the residue of a particle highly similar, if not identical, to the Higgs boson had been observed.

---

193.   Importantly, the relative plausibility system does not forbid parties from relying on multiple stories as alternative factual claims. This reliance will face an epistemic constraint: Factfinders will tend not to believe a party who tells them that A is true, but if not A, then B; and if not A or B, then C. Any such claim suggests that the party hides the truth or, at best, has no good idea what is true. Still, if a party chooses to rely on alternative stories, and factfinders determine that one of those stories wins the relative plausibility contest, then judgment should be rendered accordingly.

194.   Cheng, *supra* note 19, at 1257.

195.   *Id.* at 1278.

196.   *Id.*

197.   *See* Michael Moyer, *Have Scientists Found 2 Different Higgs Bosons?*, SCI. AM. (Dec. 14, 2012), http://blogs.scientificamerican.com/observations/2012/12/14/have-scientists-found-two-different-higgs-bosons/.

Many other disciplines similarly employ probabilistic reasoning as part of their discovery effort. Mathematics heavily influences genetics, and DNA profiling is the modern paradigmatic forensic application of statistics. Many branches of medicine, from immunology to epidemiology, employ highly sophisticated mathematical models in both discovery and application. Similar examples abound in physical chemistry, fluid dynamics, and, of course, economics as well. The list of scientific disciplines taking advantage of the power of mathematics is very long indeed, but the list is not endless. Moreover, it may well be the case that the disciplines that systematically exploit mathematics as a central methodology do not make up half of what should be included in any respectable concept of "science."[198] Here again the list is long. The realm of biology—as vast or perhaps more so than the physical sciences—uses mathematics only sporadically and in limited doses. Anthropology, astronomy, ecology, psychology, physiology, anatomy (an unquestionably physical science), neurology (another one), and (still another) chemistry all seem to have done quite well employing other research methodologies in much larger doses than applied mathematics. Some of these respected disciplines do only very limited hypothesis testing or controlled studies. As powerful as mathematics has been in the hands of the theoretical physicists, looking at "science" as a whole could easily lead one to ask the opposite of Cheng's question. Specifically, one should ask what methodologies have these organized disciplines employed so effectively that we might embrace them as tools for promoting the objectives of the law?

The short answer to this question will point to these disciplines' distinctly epistemic mode of reasoning. These disciplines formulate and examine well-articulated hypotheses featuring a coherent account of causes and effects. The validity of these hypotheses is determined by their evidential confirmations, not just by frequencies or other statistical correlations. Medical diagnoses of individual patients often follow this epistemic mode, too, as a substitution for a "one size fits all" statistic.[199]

Furthermore, when one looks more closely at the disciplines with a heavy emphasis on mathematics, two features emerge. These disciplines focus on the interactions of matter or energy describable by physical laws, or alternatively, they permit reproduction of massive and replicable frequencies that capture the relevant physical phenomena. Particle physics exhibits both of these features, which is

---

198. Ever since Popper's failure to demarcate between "science" and "pseudo-science" by invoking his (nevertheless useful) falsifiability criterion, it has become widely recognized that sharp and short definitions of "science" were elusive and potentially unhelpful as well. *See* THOMAS KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (1962); KARL POPPER, THE LOGIC OF SCIENTIFIC DISCOVERY (1934); Larry Laudan, *The Demise of the Demarcation Problem, in* 76 BOSTON STUDIES IN THE PHILOSOPHY & HISTORY OF SCIENCE 111 (R.S. Cohen & L. Laudan eds., 1983); Martin Mahner, *Demarcating Science from Nonscience, in* GENERAL PHILOSOPHY OF SCIENCE: FOCAL ISSUES 515 (Theo A.F. Kuipers ed., 2007).

199. *See* L. Jonathan Cohen, *Bayesianism versus Baconianism in the Evaluation of Medical Diagnoses*, 31 BRIT. J. PHIL. SCI. 45 (1980) (arguing that patient-specific diagnoses are generally superior to statistical ones).

precisely why it has the honor of being crowned as the "King of Sciences."[200] These features allowed the discipline to incorporate an incredibly effective mathematical analysis.

Like many "sciences," adjudicative factfinding exhibits none of these features. Our society believes in free will. Choices, not physical laws of nature, govern human affairs. The formation of those choices is inextricably complicated. The complexity in the background influences is so massive that, even if fully determined, human decision-making would look more like predicting the path of a single water molecule in fluid dynamics (a literally impossible task) than the search for the Higgs boson. As a corollary, there are virtually no stable statistics that could help courts investigate a human episode. Does a witness sweating mean that he is being evasive? Could it be that the sweating witness is actually truthful but nervous? Does failure to make eye contact mean prevarication or, alternatively, a sign of respect and good manners from a well brought-up person from a certain culture? The point is obvious. Adjudicative factfinding focuses predominantly on individual occurrences. By and large, these occurrences constitute an idiosyncratic mess, not an orderly and replicable event governed by statistical laws. Mathematical models of inference cannot help courts to make sense of these occurrences.

In general, science and law pursue fundamentally different objectives. Scientific disciplines engage in discovering, organizing, and applying hierarchical bodies of knowledge. This pursuit turns caution and rigor into the disciplines' rules of the game. The disciplines consequently develop hostility toward hasty claims that something is true. Putting things on hold is a common scientific protocol—and an attractive decision as well, given that there is normally no significant cost in postponing the delivery of the scientist's findings when her data are unclear and their implications are ambiguous.

Scientific status quo, as opposed to a legal one, also does not favor one person over another. If our courts were to operate under Cheng's implicit notion of good science, their typical decision would attest that there is insufficient data to decide what is true. The court's decision consequently would be postponed indefinitely—just as a decision as to whether the Higgs boson exists has been postponed for nearly sixty years since the hypothesis was first advanced. But justice delayed is justice denied.[201] The legal status quo virtually always favors somebody; delaying a contract, tort, or property dispute for sixty years would typically mean a victory for the defendant. Courts must decide cases one way or another without waiting for more careful and more refined studies to come out. Adjudicative factfinding is—and should be—a pragmatic quest for the best decision in the face of uncertainty.

---

200. *See, e.g.*, IWAN RHYS MORUS, WHEN PHYSICS BECAME KING (2005).

201. *See, e.g.*, Carl Reynolds, *Texas Courts 2030—Strategic Trends & Responses*, 51 S. TEX. L. REV. 951, 973 (2010) (observing that judges operate on the premise that "justice delayed is justice denied").

CONCLUSION

True to the name and spirit of both contributions discussed herein, when one proposes to redesign a foundational element of the legal system, the person bears a heavy burden of proof to show that the system is malfunctioning. A reformer must carefully review and discredit the epistemological, economic, and moral justifications that scholars have advanced in support of the current system. After all, the presumption should be that a system that has been in use for so long and that underwent multiple adjustments and refinements does not have serious operational and conceptual flaws. This presumption is rebuttable, as one should never assume that the existing system is flawless, but a reformer who undertakes to rebut the presumption ought to proceed with care and attention to detail. As sophisticated and provocative as Kaplow's and Cheng's theories are, neither meets this fundamental criterion.

Kaplow writes as though the goal of welfare optimization was alien rather than integral to the legal system, but this assumption is mistaken. As we have shown, the burden of proof doctrine operates together with other evidentiary rules and practices to promote accuracy of factfinding in individual cases. Equally important, this doctrine works in synergy with substantive liability rules to promote society's welfare. Kaplow misses these two pivotal factors, while paying little attention to the conceptual and operational difficulties of his own theory. As a result, he fails to establish that his novel mechanism of assigning liability under uncertainty will outperform extant doctrine.

Cheng writes as though the conjunction paradox is the only factor that separates the burden of proof doctrine from trial by mathematics, but this assumption is unfounded. As we have shown, our factfinding system refuses to guide itself by mathematical probability because it developed a better way of determining facts. Cheng develops a new metric that creates an alignment between extant doctrine and mathematical probability, but this alignment brings about no conceptual or operational improvements.

The burden of proof doctrine may require some refinements, but it is not broken. Contrary to Kaplow and Cheng, it is operationally sound and conceptually solid. Hence, it does not require fixing, nor least of all, a complete overhaul.